Logic in Classical and Evolutionary Games

MSc Thesis *(Afstudeerscriptie)*

written by

Stefanie Kooistra
(born April 24th, 1986 in Drachten, The Netherlands)

under the supervision of **Prof. dr. Johan van Benthem**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
September 5th, 2012	Prof. dr. Johan van Benthem
	Dr. Alexandru Baltag
	Prof. dr. Benedikt Löwe

Institute for Logic, Language and Computation

Contents

Acknowledgements

I would like to express my immense gratitude to my supervisor, Johan van Benthem, whose patient but persistent guidance gave me the motivation, support and 'know-how' I needed to learn how to write a thesis. The competences I have gained from writing my thesis with Johan are invaluable and will certainly stick with me. I would also like to thank the members of the thesis committee, Alexandru Baltag and Benedikt Löwe, for taking the time to read and evaluate my thesis. Last, I am grateful to Davide Grossi for his technical lessons, which were vital to the results in this thesis, my mentor Dick de Jongh, who in many ways guided and encouraged me throughout the process of the MoL programme, and Henk Zeevat for his time when I started my thesis.

This thesis is dedicated to my parents, Henk and Sieneke Kooistra, who in every way made it possible for me to complete this work. Thank you for your tremendous support, friendship, patience and love. Also thank you to my sister, Lydia, for being an always supportive, loving and helpful friend, never afraid to be honest and always reminding me to just "deal with it."

Bart J. Buter, I can't imagine having gone through this process without you; we took classes together, we wrote our theses in each other's company, and we faced all the related challenges and successes together. Thank you for your support, your faith in my ambitions, and patience for my sometimes eccentric style of working.

Rob Uhlhorn, thank you for always relativising everything good and bad with brilliant absurdity. Our regular debauchery kept me on track, totally straight-edge. And also thanks to Bill Mikesell, for your support, good advice, good cake, and for joining and helping us all keep things fun.

Many many thanks to my friends, family and fellow students for the many forms of support; having coffees and dinners, ranting over skype, proof–reading my thesis and more; so, *thank you* to Danielle, Maaike, Pietro, Ines, Robbert-Jan, Tjerk, Patrick, Sara, Hendrike, Silvia, Piet, Nynke, Kasper, Abel and Jelle ("finished" first, after all).

Last, this thesis will always remind me of Cody and Jimmie, my little companions, who passed away during the completion of this thesis. For the majority of my life, they were my role models for absolute contentment; never judgemental, always loving, always pondering but never thinking anything. May we all be reincarnated as cats.

Chapter 1

Introduction

Game theory is the study of a fundamental and persistent aspect of human behaviour: the strategic interaction of agents. It is a "characteristic ingredient of human culture" [4]; it can be demonstrated in many kinds of social scenarios such as bartering over the price of a squash, figuring out what film to watch when all your friends have different tastes and choosing the best word a tourist who is looking for the bathroom is most likely to understand.

Despite its application to what seems like a familiar everyday occurrence, game theory is a formal mathematical framework for social interaction; that is, "game theory uses mathematics to express its ideas formally ... however, [they] are not *inherently* mathematical" [35]. Mathematics simply gives us a formal and reliable way to define key elements of strategic interaction in game-like scenarios. In game theory, a "game" is not always a game in the recreational sense[1], but every strategic scenario is considered to have their key elements in common with "fun" games: players, actions, outcomes and the value of the outcomes for each player.

Because game theory delivers a formal and elegant model describing a very basic feature about human behaviour, game theory is deeply rooted in many social sciences. Game theory is prominent in economics, for instance. The emergence of game theory is often attributed to the 1944 publication of *Theory of Games and Economic Behaviour* by John von Neumann and Oskar Morgenstern [34], in which they represent economic and other social behaviour with many of the formal concepts game theorists use today. Other fields that use game theory are political science, linguistics, logic, artificial intelligence and psychology among others. This thesis is devoted to the study of *logic* and game theory.

The connection between logic and game theory is itself intricate. Economics and other fields are generally considered to be applications of game theory[2] but

[1] Chess, checkers, poker, Candyland ...

[2] Osborne and Rubinstein observe that"the boundary between pure and applied game theory is vague; some developments in the pure theory were motivated by issues that arose in

logic can be seen as something to apply to game theory or as a way to describe game theory.

Van Benthem, in his upcoming book *Logic in Games*, points out that "logic in games" is an ambiguous phrase; it seems to reflect the difference between logic as an application and logic as a description. Van Benthem claims that we have " 'game logics' [which] capture essential aspects of reasoning about, or inside, games" and we have " 'logic games' capturing basic reasoning activities and suggesting new ways of understanding what logic is" [4]. The former uses logic to *describe* game theory, and the latter *applies* game theory to logic. Put simply, the aim of *Logic in Games* is to describe both perspectives and the intricate connections between them. For this thesis, in general, one point van Benthem makes is salient: "Some students ... [prefer] one direction while ignoring the other" [4]. To some degree, this preference likely occurs for some students by taste or inclination, and, accordingly, this thesis adheres to the preference for *game logics*.

The goal of this thesis is to take initial but key steps towards the inclusion of *evolutionary* game theory in that game logic debate. That is, it explores how we can describe evolutionary game theory by means of logic. Evolutionary game theory is an expansion of game theory into biology, where it is fitted to observe the stability of behaviour of populations of players in the animal kingdom over time. Moreover, it can also be used to understand the dynamics of our human behaviour where classical game theory would otherwise fall short.

Although the step from classical to evolutionary game theory looks small (it is "simply" thinking of games in terms of populations instead of individuals, and it is still mathematically described by all the same components), it revolves around a dramatic change of perspective on the traditional mathematical components of game theory. But given that the study of *classical* game logics is already firmly grounded[3], one may conclude that game logics should be easily extended to evolutionary game theory since the theories seem so closely related.

John Maynard Smith, one of the earliest authorities on evolutionary game theory, claimed the following in his article "Evolutionary Game Theory" [41]:

applications. Nevertheless, we believe that such a line can be drawn ... [This book] ... stay[s] almost entirely in the territory of 'pure' theory" [35]

[3]The Institute for Logic, Language, and Computation (ILLC) at the Universiteit van Amsterdam is one of the few institutions with a program specifically about logic and games. It is (or has once been) home to many of the academics who have given shape to the study of game logics: Johan van Benthem with many innovating publications on the subject as well as peripheral topics, Eric Pacuit, Alexandru Baltag, Benedikt Löwe, Olivier Roy, Boudewijn de Bruin, Wiebe van der Hoek, Marc Pauly, Robert van Rooij and more. One could, of course, go on to mention many more distinguished contributors to this field at other institutions around the world.

There are two main differences between classical and evolutionary game theory.

1. *The replacement of "utility" by "fitness"...*
2. *The replacement of rationality by natural selection.*

This thesis will address and elaborate upon Maynard Smith's points. The following two issues will therefore be major themes throughout this thesis, and will be the main obstacles to overcome in establishing an evolutionary game logic.

1. Many established game logics are not sufficient to simply extend to evolutionary game theory. Whereas classical game theory is elegantly described by modal logics with modalities for knowledge and preference, evolutionary game theory raises two classes of issues. First, what the logic expresses and what evolutionary concepts mean do not always coincide. Second, evolutionary game theoretic concepts are formally different to such a degree that it implies finding a more expressive language.

   In order to figure out how to remedy these issues, we take a close look at the essential differences between classical game theory and evolutionary game theory, how those differences influence the logical structure of classical and evolutionary game theory, and then search in our logical toolbox for a solution.

2. Rationality, which underpins the enterprise of game theory, is radically questioned in the evolutionary game setting. The nature of evolutionary game theory is such that the players are programmed to play strategies instead of burdened to rationalize and deliberate over what they should *choose*. It is therefore prudent to ask, what is game theory without rationality and choice?

This thesis will accomplish the above goals by describing the basic background and details of classical and evolutionary game theory, exploring some appealing logics for classical game theory and then exploring how those fit with evolutionary game theory.

Chapter 2 introduces the basic mathematical concepts and definitions of classical game theory. As mentioned above, game theory is mathematical and is therefore the tool used to elegantly and consistently describe the components of a game; players, actions, outcomes, and preferences. Classical game theory has two forms *strategic* games and *extensive* games. A strategic form game describes a one-time simultaneous strategic event between to or more players, and an extensive game describes a turn-taking sequence of events. This chapter will introduce the formal terminology of classical game theory as well as what we can say about game with them. This includes the Nash equilibrium and various other solution concepts. This chapter will also emphasize the rationality assumptions that are foundational to the solution concepts.

Chapter 3 introduces evolutionary game theory. I will motivate the evolutionary perspective by explaining its unique background with its historical origins in Darwin's expedition on the Beagle. Understanding the origins of evolutionary game theory will foreshadow not only why it is very interesting to look at in logic, but also the challenges we will face given what it stands for. This chapter will describe two basic formal approaches to evolutionary game theory: evolutionary stability and replicator dynamics. It will also take the opportunity to comment on the role that rationality plays in evolutionary game theory.

The crucial chapter 4 takes a detailed look at how classical game theory in strategic and extensive form has been and can be described by existing logics. An in–depth analysis of classical game logic is *paramount* to motivating the evolutionary game logic in the following chapter. Game theory and logic are akin, for they share many formal components such as possible worlds or states and reasoning agents with preferences. This chapter will focus primarily on examples of static and dynamic modal logics as evidence showing how game theory resembles logic. A result of this analysis is a set of relational modalities *freedom*, *knowledge*, and *preference*, which express key relationships between the states in a strategic form game. In combination with *hybrid logic*, a logic that "brings to modal logic the classical concepts of identity and reference" [12], it is possible to elegantly redefine some notions in the above modal logic for both strategic and extensive form games; this new perspective on those notions consequently sheds some light on their meaning.

Chapter 5 finally approaches the question of whether the logic–game theory interface of chapter 4 is applicable given what we know about evolutionary game theory from chapter 3. I will argue that the difference in interpretation of classical game theoretic terms in evolutionary game theory is responsible for the issues arising in inventing a logic for evolutionary games. My approach to this investigation is two–fold: the reinterpretation of terms has effected the role that rationality plays in evolutionary games, and the reinterpretation of terms has also effected how we must think of the logical components of game theory such as players, strategies and preferences:

First, there is a significant disparity between what rationality means for classical game theory and evolutionary game theory, and that has consequences for the logic of classical and evolutionary game theory. I consider three ways we can "think of" rationality that may fit evolutionary game theory and logic of evolutionary game theory without compromising the evolutionary perspective.

Second, because the game theoretic components players, strategies and preferences, have been reinterpreted under the evolutionary perspective, we should also reinterpret them logically for a correct evolutionary game logic. I propose two ways of doing this; first by means of the hybrid logic with freedom, knowledge and preference as binders, and second by means of an introduction of a new relation simply based defining strategies as players. Defining evolutionary game theoretic terms in the latter logic will prove to be the most intuitive alternative, for it best expresses what evolutionary game theory stands for.

In conclusion, chapter 6 summarizes everything that has been discussed in this thesis towards including evolutionary game theory in the game logics debate. It describes the novel results appearing in this thesis and identifies possible next steps and areas of development. The topics and results described are complex and highly connected within the fields of logic and game theory, so there are many facets of the intersection that remain to be explored. I suggest some future work towards this end as well.

Chapter 2

Classical Game Theory

Classical game theory establishes the framework for studying and analysing strategic interaction. Strategic interaction incorporates the decision-making behaviour of rational intelligent agents with preferences over the possible outcomes, which are determined by the joint actions of all agents. In general, the study of classical game theory specifies terminology, the solution concepts and the reasoning behind the processes of strategic interaction. Classical game theory provides the theoretical background and structure on which many practical studies of strategic interactions base their investigations. Classical game theory is very popular and informative to researchers in various fields in which strategy is involved.

There are two forms of classical games: the strategic form game and the extensive form game. The former describes games in which players act once and simultaneously. This form is straightforward; players can only deliberate beforehand, act, and then accept what follows from his and his opponents' actions. An instance of a game that takes this form is rocks, paper, scissors. Neither player is to observe his opponents action before taking his own. Therefore, a player could only choose what to play based on what he thinks his opponent *might* play.

The latter, the extensive form game, is more complex. It describe strategic interaction that takes a sequential, or turn taking, form. A game such as chess is an instance of an extensive form game. The players take turns one after the other, until one player reaches a last move which concludes the game at an outcome in which he wins or loses. This chapter will describe the terminology and results of both forms.

This chapter will focus on information about strategic and extensive form games that will be relevant to the upcoming discussions in this thesis. The interaction of players, actions and preferences are crucial to eventually coining a logic that can describe classical games, so that will be emphasized in this chapter. Moreover, this chapter is to equip the reader with knowledge of the features of classical game theory that will, in chapter 3, be adapted under the development on classical game theory, evolutionary game theory. Evolutionary game theory has the same basic machinery as classical game theory, but the meaning of the terminology is changed. Because these new meanings imply that, among other things, rationality is no longer a factor, I will describe the influence that rationality has on classical game theory. By being aware of that influence, we will be able to evaluate the consequences of abandoning rationality in theories based on classical game theory, such as evolutionary game theory.

2.1 Strategic Form Games

A strategic form game is the framework by which game theorists express one-shot strategic interaction. The main factors that compose this framework are players, actions, and preferences over the outcomes that result from the joint actions of the players. This section will first describe the basic terminology, a tool that visualizes the main idea of the theory and some crucial solution concepts and results in classical game theory.

2.1.1 Terminology

A player is a decision–maker who is to choose an action in a game. *Agent* is a more specific term referring to a player who we assume to be intelligent and rational. A strategic form game is defined[1] as:

Definition 2.1.1. *A* ***strategic form game*** Γ *is a tuple* $\langle N, (A_i)_{i\in N}, (\succeq_i)_{i\in N}\rangle$ *where:*

> N *is the set of players.*
>
> $(A_i)_{i\in N}$ *is a non-empty set of actions for each player* $i \in N$
>
> $(\succeq_i)_{i\in N}$ *is the preference relation for player* $i \in N$ *on the set of action profiles* $A = \times_{j\in N} A_j$

This is the basic definition of a strategic form game. An action profile $a \in A$ represents the list of actions $(a_1, ..., a_n)$ played by each player $i \in N$, where a_i is the i–th projection. An $a \in A$ is also an *outcome* of the game, for it denotes the unique situation resulting from each player's action.

The preference component of a strategic form game can be further specified by the strict and weak variety:

- $a \succeq_i a'$ if i prefers a more than or equally to a'
- $a \succ_i a'$ if $a \succeq_i a'$ and $a' \not\succeq_i a$
- $a \simeq_i a'$ if $a \succeq_i a'$ and $a' \succeq_i a$

[1]Components of this definition originate in [35] and [10]

Preference over outcomes can also be expressed in terms of utility. Utility expresses that an outcome has a particular value to a player, which is determined by a numerical consequence of that outcome. Preference and utility achieve the same goal, but by different means. The choice to express a game by utility or preference is usually a question of convenience, i.e. sometimes it is just more appropriate to talk about preference than utility or vice versa. So, in this discussion, as well as others, both concepts will be used where appropriate. The formal guide to connecting utility with preference is[2]:

Definition 2.1.2. *$U_i : A \to \mathbb{R}$ defines a preference relation $\succeq_i$ by the condition that $a \succeq_i a'$ if and only if $U_i(a) \geq U_i(a')$.*

A strategic form game may also represent scenarios where players take mixed actions. For instance, player i chooses to play action a_i p percent of the time and a'_i p' percent of the time[3], where $\sum_{a_i \in A_i} p(a_i) = 1$ and for all $a_i \in A_i$, $p(a_i) \geq 0$. An alternative possible interpretation of a mixed strategy is as the probability by which a player i may play an action in A_i. The definition of a strategic form game that accounts for mixed actions is:

Definition 2.1.3. *The **mixed** extension of the **strategic form game** Γ is a tuple $\langle N, \Delta(A_i)_{i \in N}, (U_i)_{i \in N} \rangle$ where:*

$\Delta(A_i)_{i \in N}$ is the set of probability distributions over A_i.

$(U_i)_{i \in N}$ is a function from mixed action profiles to real numbers, $(U_i)_{i \in N} : \times_{j \in N} \Delta(A_j) \to \mathbb{R}$.

Strategy is the default term that describes actions of players, in *general*, whether they are playing a pure action or mixed set of actions. From now on, actions of a player i will be denoted as strategies σ_i even though it is usually a pure strategy where a_i would be sufficient[4]. We thus identify outcomes with the set $S = \times_{i \in N} A_i$ of *strategy* profiles [10].

Definition 2.1.4. *A **strategy profile** $\sigma \in S$ denotes the list of pure or mixed actions played by each player $i \in N$, $\sigma = (a_1, ..., a_n)$, where σ_i denotes the ith projection, i.e. $\sigma_i = a_i$ and σ_{-i} denotes the choices of all players except i, $\sigma_{-i} = (a_1, ..., a_{i-1}, a_{i+1}, ..., a_n)$. [10]*

This concludes the basic terms that compose strategic form games. In the following section, concepts such as the game matrix, solution concepts and procedures will be introduced.

[2]Based on [35].

[3]This interpretation of a mixed strategy alludes to multiple plays of a game.

[4]Strategic form games do not exhibit the complexity seen in other game forms, and therefore the difference between the terms actions and strategy are of little consequence for now. Nevertheless, we wish to distinguish between actions and strategies in preparation for other models relevant to this thesis

2.1.2 Game Matrix

A game matrix is a visual tool that lists one player's actions in a row and the other's in a column, which results in a grid or table. The box where a column and row converge represents a strategy profile and is labelled by the utilities for each player assigned to that strategy profile. Although a game matrix is limited to representing only games with two players, it is a useful tool that can efficiently describe and visualize many fundamental concepts. The following figure is an arbitrary example of a game matrix, where player i's actions a_i and b_i are listed in the left column, player j's actions a_j and b_j are listed in the top row, and the cells represents the four possible combinations of i's and j's strategies:

	a_j	b_j
a_i	$u_i(a_i, a_j)$, $u_j(a_i, a_j)$	$u_i(a_i, b_j)$, $u_j(a_i, b_j)$
b_i	$u_i(b_i, a_j)$, $u_j(b_i, a_j)$	$u_i(b_i, b_j)$, $u_j(b_i, b_j)$

Each cell in the game matrix represents the outcome of the game from the strategy profile of the players' strategies at the cell's respective row and column. It is prudent to note that "one often identifies the outcomes with the set of *strategy profiles* ..." [4], and in the chapters about game theory and logic, we will see that "it seems natural to use the strategy profiles themselves as possible worlds" [4].

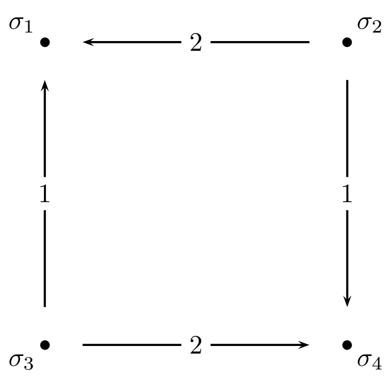


Figure 2.1: This picture represents the four states and a few preference relations.

Because games are often described in terms of preference instead of utility, it is also possible to construct a corresponding "preference matrix," which is a pointed preference model. Instead of outcomes represented by cells, the outcomes are represented by points (possible worlds or states), and the preference is indicated by an arrow pointing towards a preferred world (away from every world that is less preferred). Figure 2.1 depicts an example of a preference model with some of the players' preferences between worlds marked.

Types of Games

By observing relationships between the outcomes of a game, we can also make specific claims about properties of a game; that is, we can distinguish between types of games. These include symmetric, zero-sum, and coordination games.

Symmetric Game
A symmetric game implies that in the game matrix, the payoffs and corresponding utilities of one player are the transpose of the pay-offs/utilities of the other, and each player has the same strategies available to him.

	a_{-i}	a'_{-i}
a_i	(n, n)	(k, l)
a'_i	(l, k)	(m, m)

Notice that the rule holds that a 2-player game Γ is symmetric if $\forall ab : u_1(a, b) = u_2(b, a) = u(a, b)$. Because it does not matter which player is at the row or column position, it is sufficient to talk about utility in terms of just the row player $u(a, b)$.

Coordination Game
A coordination game is a type of game where it is the best for both players to choose the same strategy. In general, a coordination game has the following game matrix composition:

	A	B
A	N, n	L, l
B	M, m	K, k

where $N > M$, $K > L$, and $n > l$, $k > m$.

Thus, either case of choosing the same strategy has a higher utility than either case of choosing differing strategies.

Zero-Sum Game
A zero-sum game is strictly competitive; the gain of one player is the loss of another. An example of a zero-sum game is the Matching Pennies game: each player flips a coin, and the row player wants the pennies to match (both heads of both tails), and the column player wants the pennies to be opposites (one heads, one tails). The matrix is composed as:

	Head	Tail
Head	1, -1	-1, 1
Tail	- 1, 1	1, -1

We see that what one player's positive utility, is the other's negative utility. This is characterised by the rule for zero-sum games that for any strategies A, B in a game, $u_1(A, B) + u_2(A, B) = 0$. Therefore, it is called a *zero-sum* game.

The interaction between players in a strategic scenario is described by the definition and visualised by game matrix, but it does not yet address the process involved between the player and his action; that is, what motivates players to choose certain actions? Essentially, a player will choose an action that is in his best interest.

Analyses of how players may reason towards his most preferred outcome given his opponents' moves have resulted in concepts such as the Nash equilibrium (NE), which is one out of many possible solution concepts in game theory. The following section describes the some of most central results and solution concepts to classical game theory.

2.1.3 Solution Concepts

A game theorist can reason about a game and decide which strategies rational and intelligent players in a game will play. The resulting strategy profiles can be given names and defined on the basis of the reasoning for why a game will result in that strategy profile.

John Nash's Ph.D. thesis published in 1950 [33], which followed von Neumann and Morgenstern's *The Theory of Games and Economic Behaviour* [34] by only six years, introduced what is arguably the most important solution concept in game theory, the Nash Equilibrium.

Definition 2.1.5. *A strategy profile σ^* is a **Nash Equilibrium** if for every player $i \in N$*

$$(\sigma_i^*, \sigma_{-i}^*) \succeq_i (\sigma_i, \sigma_{-i}^*) \text{ for all } \sigma_i \in S_i$$

Nash equilibrium is also measurable by another means; the best response function. This function determines which action is the best *response* to an opponent's strategy, and is defined [35] as:

Definition 2.1.6. *A strategy σ_i is a **best response** to $-i$'s strategy, $B_i(\sigma_{-i})$, if*

$$(\sigma_i, \sigma_{-i}) \succeq_i (\sigma_i', \sigma_{-i}) \text{ for all } \sigma_i' \in S_i$$

Note that a the best response function is not always one-to-one. If there is a second (or more) strategy with the same utility as a best response, then it is also a best response. Therefore, the best response function results in a set of strategies. The set may be a singleton, but when it is not, it follows that there exist multiple equally preferred best responses. If both players are playing a best response to their opponent's best response, the result is a Nash equilibrium. The Nash equilibrium in terms of best response is thus defined as [35]:

Definition 2.1.7. *A strategy profile σ^* is a* ***Nash Equilibrium*** *if for every player $i \in N$.*

$$B_i(\sigma^*_{-i}) = \sigma^*_i$$

Nash proves in [33] that "every finite game has an equilibrium point." He demonstrates that if a game does not have a pure strategy Nash equilibrium, then there exists a Nash equilibrium in mixed strategies[5].

The following game is a famous and authoritative example of a strategic game and demonstrates the Nash equilibrium solution concept.

Example: The Prisoner's Dilemma

This game describes a scenario where two individuals are arrested as suspects to a crime. Because they have insufficient evidence to convict either suspect, the arresting policemen offer the suspects the same deal: if the suspect cooperates and thus tattles on the other and the other defects and remains silent, the suspect who tattled goes free. If they both tattle on each other, both go to jail for 5 years. If they both remain silent, they each get jailed for 1 year. They must each therefore choose to *cooperate* or *defect.* This is represented in the following matrix:

	Cooperate	*Defect*
Cooperate	3, 3	0, 6
Defect	6, 0	1, 1

In the Prisoner's dilemma the outcome (*Defect, Defect*) is a Nash equilibrium. If the row player changes his action, he will receive a lesser utility of 0, and if the column player changes his strategy, he will also receive a lesser utility of 0.

Nash equilibrium, as well as other solution concepts, have refinements. The Nash equilibrium described in definition 2.1.5, also referred to as a weak Nash equilibrium, has a refinement where the players strictly prefer one outcome over the other. This is called the strict Nash equilibrium:

Definition 2.1.8. *An outcome σ^* is a* ***Strict Nash Equilibrium*** *if for all $\sigma_i \in S_i$,*

$$(\sigma^*_i, \sigma^*_{-i}) \succeq_i (\sigma_i, \sigma^*_{-i}) \text{ and } (\sigma^*_i, \sigma^*_{-i}) \not\preceq_i (\sigma_i, \sigma^*_{-i})$$

The following example demonstrates some additional refinements of Nash equilibrium.

[5]To prove this, Nash relies on Brouwer's fixed point theorem. The proof consists of a mapping that satisfies the conditions of Brouwer's fixed point theorem; that it is compact, convex, and closed. This therefore requires that the function has fixed points. Nash proves that these fixed points are exactly the equilibrium points [32].

Example: Stag Hunt Game

There are many more refinements of the Nash equilibrium that reveal interesting additional information about a game when there are multiple Nash equilibria. Consider the following example of the Stag Hunt game, which has multiple Nash equilibria[6]:

	Stag	*Hare*
Stag	2,2	0,1
Hare	1,0	1,1

(*Stag, Stag*) and and (*Hare, Hare*) are the Nash equilibria in this coordination game. Two refinements of Nash equilibrium are exemplified in this game: pay–off dominant and risk–dominant equilibria. (*Stag, Stag*) is the pay–off dominant equilibrium, because it yields the highest utilities, and (*Hare, Hare*) is the risk–dominant equilibrium, because it has the lowest risk of a low utility should one player deviate for some reason.

2.1.4 Iterated Elimination of Dominated Strategies

A special procedure of reasoning about a game is by means of *iterated elimination of dominated strategies* (IEDS), which is the repeated application of the notion *dominated strategy* to a game. With each application of IEDS a dominated strategy is removed from the game.

It operates under the assumption that a rational player will never choose to play strategies that give him unilaterally (for *any* strategy the opponent plays) strictly lower utilities than at least one other strategy. Following this reasoning, one may repeat this until a smaller game or only one outcome remains. This procedure depends on the solution concept dominated strategy:

Definition 2.1.9. *A strategy σ_i is a **dominated strategy** for player i if*

$$\exists \sigma_i' \forall \sigma_{-i} (\sigma_i', \sigma_{-i}) \succeq_i (\sigma_i, \sigma_{-i})$$

In other words, if there is a better strategy σ' than the one in question σ for every possible move of i's opponent, then σ is a dominated strategy. A rational player j will reason that his opponent i will never choose to play a dominated strategy, so j may rule out the outcomes where i plays a dominated strategy. Player i will reason that because player j has ruled out one of his own dominated strategies, one of j's strategies becomes dominated, and can be eliminated as well. This is repeated as many times as it takes for no more strategies to be deleted. The process will result in a strategy profile (one or more) that survive after the sequence of deletions.

Consider the game in the following example and the single outcome, (D, A) that the IEDS procedure selects.

[6]The story of the Stag Hunt game is as follows: two hunters can choose each to hunt for hare or stag. Hunting hare is safer but less rewarding than hunting a stag. The stag is far more rewarding, but if a hunter hunts a stag alone, he is likely to get injured. Thus, if the two hunters work together to catch a stag, they will split the large reward, which is better than each catching and not sharing a hare.

	A	B	C
D	2,3	2,2	1,1
E	0,2	4,0	1,0
F	0,1	1,4	2,0

Figure 2.2: Matrix 1 borrowed from [2]

The IEDS procedure in this example is described in the following figure 2.3:

$\Rightarrow$

	A	B
D	2,3	2,2
E	0,2	4,0
F	0,1	1,4

$\Rightarrow$

	A	B
D	2,3	2,2
E	0,2	4,0

$\Rightarrow$

	A
D	2,3
E	0,2

$\Rightarrow$

$\star$	A
D	2,3

Figure 2.3: Matrices (2), (3), (4), ($\star$) respectively

In matrix **(1)**, the initial game, the strategy C for the column player is dominated resulting in matrix **(2)**. Here we can see that the strategy F is now dominated for the row player, which, after deletion, results in matrix **(3)**. Following this procedure, we can delete strategy B resulting in matrix **(4)** and then E resulting in the solution matrix $\star$ where (D, A) is the solution that survives the process of IEDS.

2.1.5 Conclusion

It is prudent to note that we are bound to strict assumptions of rationality and reasoning ability in order to make solution concepts like Nash equilibrium, best response and IEDS realistic. If we did not assume that the players consistently reasoned towards and chose the strategies that would result in the most preferred outcome possible, then Nash equilibrium, IEDS and best response would be hard to justify. These solution concepts all require rationality in some form[7].

The strategic form is nevertheless an intuitive and robust way to describe strategic scenarios, but this form only partly composes classical game theory. The extensive form game is, similarly to the strategic form game, a model by which we can understand the strategic interaction and adheres to the stringent rationality requirements as well.

An initial definition of an extensive form game by von Neumann in 1928 was further developed in 1953 by Harold Kuhn, a long-time colleague and friend of Nash, into what is now the default definition of an extensive form game. As we will see in the next section, the extensive form game is expressive like the

[7] See [37] for an in-depth account of the rational requirements for rationality in (extensive) games.

strategic form game, but it models features of games that the strategic form is not equipped to handle, namely sequential turn-taking games.

2.2 Extensive Form Games

Whereas a strategic form game expresses a unique one-time, one-move scenario, an extensive form game can express a game that involves sequences of actions and turn taking. As with strategic form games, an extensive form game is defined by players, actions and utilities of outcomes, but there are some differences that clearly sets it apart from a strategic form game. For instance, an extensive form game is represented not by a matrix but by a *game tree*, it introduces a new notion *history*, and it uses an alternative perspective on the notion *strategy*. With these concepts, an extensive form game represents all possible sequences of the players' choices, which lead to unique outcomes that are associated with one particular sequence. Those outcomes are also assigned utilities for each player in the game.

In this section, I will first describe the basic terms and definitions of extensive form games. Second, I will describe the game tree. Also, I describe solution concepts for extensive form games. This includes the Nash equilibria which exist for extensive form games, but because they ignore the sequential nature of the game, another solution concept, *subgame perfect equilibrium* is used to more accurately describe a solution intuitive to the game.

Generally, extensive form games operate under the assumption that the players observe each other's moves and know what state they are in; under these circumstances the game is a *perfect information* extensive form game. However, there are extensive form games with *imperfect information* meaning that the players are not always aware of the other player's move. I will describe these games as well.

2.2.1 Terminology

The following terms and definitions are used to describe extensive form games[8]. We formally define an extensive form game as

Definition 2.2.1. *An* ***extensive form game*** *is a tuple*
$G = \langle N, H, P, (\succeq_i)_{i \in N} \rangle$ *where:*

Players As with strategic form games, there is a set of players N, with $i \in N$ as some individual player.

Histories A set of histories H describes all possible sequences of actions taken by the players, where $h = \emptyset$ is the start of the game (i.e. the empty sequence). If a member of H (a sequence of actions) is $h^K = (a_1, ..., a_K)$, then for every $L < K$, it holds that $h^L = (a_1, ..., a_L)$ is a member of H. A history $h^K = (a_1, ..., a_K)$

[8]Based on [35]

is a member of the set of terminal histories[9] $Z \subset H$ if there is no a^{K+1} such that $h^{K+1} = (a_1, ..., a_{K+1})$ is in H. The game ends at the terminal history, but after a non-terminal history h, the set of actions available are denoted by $A(h) = \{a : (h, a) \in H\}$.

Turns Each non-terminal history $h \in H \backslash Z$ is assigned to a player in N by the player function P. Thus $P(h) = i$ indicates that after h, it is player i' turn.

Preferences For each player i there is a preference relation $(\succeq_i)$ over the set of terminal histories Z. We denote $(z, z') \in (\succeq_i)$ as meaning that i prefers z over z' or prefers them equally.

In an extensive form game, a strategy has a more involved notion of strategy than in strategic form games:

A strategy, so far, has been understood to be the action or probability distribution of actions that a player can take in a game. In extensive form games, however, players take multiple actions progressing through the game, so we take an alternative approach to account for this. A strategy in extensive form games specifies for each history which action will be taken by the player whose turn it is. That is, a player has a planned response to every action that his opponents may take. Formally, a strategy for i associates with each h for which $P(h) = i$ an action $a \in A(h)$.

Because a strategy is a sequence of actions, an outcome cannot be described directly by strategy profiles. Instead we can define an outcome for every strategy profile σ as $O(\sigma)$ "is the terminal history that results when each player $i \in N$ follows the precepts of σ_i" [35]. This only holds for perfect information games.

2.2.2 Game Tree

A game tree is a non-cyclical pointed graph that is a visual tool representing extensive form games as described above. Like the matrix for strategic form games, the game tree is useful to understand the concepts described by the formal terms. The extensive form game, and thus the game tree, differs from the game matrix, because the game matrix "...describes only a situation where each player makes a single choice, in ignorance of the choices made by the other players, and the game is then over. The tree thus appears to be a more general description, allowing players to move more than once and also to observe what other players do" [14].

The game tree is composed of nodes, which represent histories, and edges which represent possible actions a player can take given the preceding node. The terminal nodes are also labelled by utilities, which correspond to the preference ordering over outcomes for each player[10] These concepts are illustrated by the tree in figure 2.4.

[9] I exclude the possibility of infinite histories, for it is not currently relevant.

[10] The utilities can alternatively be represented as preferences. See section 2.1.1.

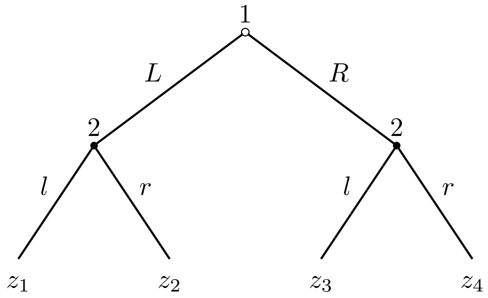


Figure 2.4: This is an extensive form game with the terminal nodes marked by zs.

2.2.3 Equilibria

Extensive form games have as a solution concept Nash equilibrium, but it also introduces the solution concept subgame perfect equilibrium. Nash equilibrium [35] in extensive form games essentially claims the same thing as in strategic form games. No player can do better by defecting from his current strategy.

Definition 2.2.2. *A* ***Nash equilibrium of an extensive game with perfect information*** $\langle N, H, P(\succeq_i)\rangle$ *is a strategy profile* σ^* *such that for every player* $i \in N$ *we have*

$$O(\sigma_i^*, \sigma_{-i}^*) \succeq_i O(\sigma_i, \sigma_{-i}^*) \text{ for every strategy } \sigma_i \text{ of player } i$$

However, Nash equilibrium does not reflect the sequential nature of extensive form games, and the following example, borrowed from [35], verifies this and consequentially motivates a solution concept reflecting the sequential nature of extensive form games, the subgame perfect equilibrium.

The following extensive form game, borrowed from [35], exhibits how the solution concept Nash equilibrium can lead to unintended conclusions.

The game in figure 2.5 has two Nash equilibria: (A, r) and (B, l) . However, the equilibrium (B, l) is motivated by the explanation that player 1 will play B because of player 2's "threat" of playing l. Playing l still amounts to the same payoff, after all. But at player 2's choice node it would never be optimal to play l because playing r affords a greater payoff, so the threat of playing l is incredible. Therefore, it is better to base an equilibrium for extensive form games on what is optimal for an acting player at every node h. This relies on the notion *subgame*. The definition of a subgame of an extensive form game Γ with perfect information is [35]:

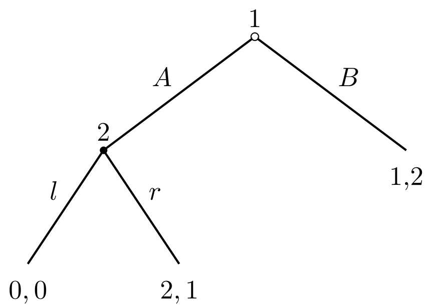


Figure 2.5

Definition 2.2.3. *A* ***subgame*** *of the extensive game with perfect information* $\Gamma = \langle N, H, P, (\succeq_i)\rangle$ *that follows history* h *is the extensive game* $\Gamma(h) = \langle N, H|_h, P|_h, (\succeq_i |_h)\rangle$ *where* $H|_h$ *is the set of sequences* h' *of actions for which* $(h, h') \in H$*,* P_h *is defined by* $P|_h(h') = P(h, h')$ *for each* $h' \in H|_h$*, and* $\succeq_i |_h$ *is defined by* $h' \succeq_i |_h h''$ *if and only if* $(h, h') \succeq_i (h, h'')$*.*

The subgame perfect equilibrium is a solution where after each history h, each player's strategy is the optimal one given his opponent's move. By basing an equilibrium concept on subgames, we preserve the sequential nature of the scenario. Thus, a subgame perfect equilibrium is based on the concept of a subgame:

Definition 2.2.4. *A* ***subgame perfect equilibrium of an extensive game with perfect information*** $\Gamma = \langle N, H, P, (\succeq_i)\rangle$ *is a strategy profile* σ^* *such that for every player* $i \in N$ *and every non-terminal history* $h \in H \backslash Z$ *for which* $P(h) = i$ *we have:*

$$O_h(\sigma^*|_h, \sigma^*_{-i}|_h) \succeq_i |_h O_h(\sigma_i, \sigma^*_{-i}|_h) \text{ for every strategy } \sigma_i \text{ of player } i \text{ in the subgame } \Gamma(h)$$

Where O_h *is the outcome function of* $\Gamma(h)$*.*

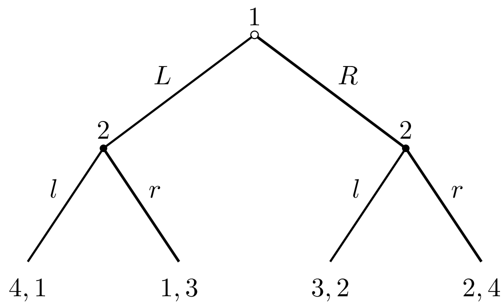
It is also possible to view extensive form games as a strategic form game by means of normalisation, and subsequently we can examine solutions in strategic form.

2.2.4 Normalisation

Because we can list all the strategies available to each player, it follows that we can convert any extensive form game into a strategic form game by putting the strategies into the matrix and assigning the corresponding utilities. Formally [35],

Definition 2.2.5. *The* ***strategic form of the extensive game with perfect information*** $\Gamma = \langle N, H, P, (\succeq_i)\rangle$ *is the strategic game* $\langle N, (S_i), (\succeq'_i)\rangle$ *in which for each player* $i \in N$

- S_i *is the set of strategies of player* i *in* Γ
- $\succeq'_i$ *is defined by* $\sigma \succeq'_i \sigma'$ *if and only if* $O(\sigma) \succeq_i O(\sigma')$ *for every* $\sigma \in \times_{i\in N} S_i$.



(a) Extensive form game tree.

	L	R
ll	(4,1)	(3,2)
lr	(4,1)	(2,4)
rl	(1,3)	(3,2)
rr	(1,3)	(2,4)

(b) Strategic form game matrix.

Figure 2.6: (b) is the strategic form game matrix that results from the extensive form game tree in (a)

Notice in figure 2.6 that expressing an extensive form games in this manner is inefficient, for the conversion can lead to very large matrices. This process leads to redundantly listing outcomes. Notice in the matrix above that the outcome with utilities (4,1) is listed twice (as are all the outcomes) for the strategy profiles (L, ll) and (L, lr) are the same.

Given player 1 plays L, both of the strategies ll and lr express that he plays l when 1 plays L and l when 1 plays R . Because the strategy profile already specifies that 1 plays L, the fact that 2 may play l or 1 after 1 plays R has no bearing on the outcome. Therefore, expressing extensive form games in matrices leads to inefficient redundancies. The comparison does reflect an interesting fact about matrices:

> The reason is that the matrix can, in fact, be thought of as modelling any interaction, even ones in which the players move more than once. The key idea here is that strategies of the matrix can be thought of,

not as single moves, but rather as complete plans of action for the tree. [14]

The fact that this is possible will be of interest in the upcoming discussion on logic in classical game theory.

2.2.5 Backward Induction

The procedure of backward induction is the process that proves Kuhn's Theorem that "every finite extensive game with perfect information has a subgame perfect equilibrium" [35]. The idea is that there is a best terminal node branching from a node h for the player whose turn it is. If that player is rational, he will choose the action corresponding to that terminal node. For this reason, we can back that value up the tree to that node h. If the player whose turn it is at the node leading to this newly valued node h, and that value is preferred to his other actions, then he will take that action, allowing us to move that value further up the tree. Continuing in this fashion, we can determine a path in the tree that selects an outcome called the backwards induction solution. For an in-depth discussion of backwards induction see sources such as [5], [9] and [26].

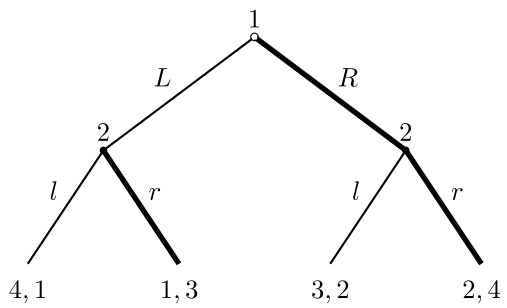


Figure 2.7

The bold lines in the game in 2.7 represent the path in the game that results in a backward induction solution. Every game has a unique backward induction solution if the preference over terminal histories is strict for each player. Otherwise, the backward induction path would split at any point h where the values of the possible actions are the same. Both outcomes resulting from h would be backwards induction solution.

2.2.6 Imperfect Information

An imperfect information extensive form game is an extension of extensive games where the players at some point in the game do not have knowledge of past moves. This implies that a player, at some point in the game, is uncertain of which state he is in. This occurs when his opponent's move is not public to him, or if moves are made by chance. These games have added structure that represent what knowledge the acting players have about the state of the game. Furthermore, the special player chance is added. The definition accounts for these factors with two additional concepts in the tuple:

Definition 2.2.6. *An* ***imperfect information extensive form game*** *is a tuple* $G = \langle N, H, P, f_c, (\mathcal{I}_i)_{i \in N}(\succeq_i) \rangle$*, where:*

N, H, P, and $\succeq_i$ These are as described in definition 2.2.1 above.

Probability Distribution The actions of player chance are determined by the probability distribution $f_c(\cdot|h)$ on $A(h)$. That is, the probability that player chance plays a at history h is $f_c(a|h)$. Because of the chance component, the utility function is defined as lotteries over terminal histories, since chance induces a non-deterministic component over terminal histories.

Information Partition The structures that represent the knowledge of the acting players are information sets. Information sets partition the set of all histories where a player acts i.e. $\mathcal{I}_i$ is a partition on $\{h \in H : P(h) = i\}$, and an information set is a $I_i \in \mathcal{I}_i$. These structures are interpreted as follows, the acting player i cannot distinguish between h and h' when $h, h' \in \mathcal{I}_i$.

Because players cannot distinguish between histories in the same information set, they have the same set of actions to choose from if the histories are members of the same information set. Hence $A(h) = A(h')$ whenever $h, h' \in \mathcal{I}_i$. For if $A(h) \neq A(h')$ players could deduce in what history they are by the actions available to them.

Furthermore, because players cannot distinguish between histories in the same information set, they can only decide upon one (stochastic) action per information set. Therefore, histories are no longer the primitive of the game over which is reasoned, but rather information sets take its place.

In general, backward induction is not possible under imperfect information [26], because in imperfect information games a player chooses an action per information set. In backward induction, you assign a value to every history. Moreover, a player takes an action per history. However, in imperfect information games, a player chooses an action per information set. Therefore, it is unclear how to assign values to actions in the information set. Backward induction is generally not possible for imperfect information games[11].

2.3 Conclusion

With these basic concepts that compose classical game theory, we can move on to describe evolutionary game theory. There are multiple ways one can choose

[11] Some new literature suggests that for specific imperfect information games backward induction-like algorithms are possible [15].

to think about evolutionary game theory in relation to classical game theory. Because evolutionary game theory takes terms and solution concepts and applies it to patterns in biology, one could argue that it is simply an application of game theory. Evolutionary game theory also reveals patterns about the behaviour of collections of individuals and invents new solution concepts, so it can also be considered to build on classical game theory as a development or extension.

Classical game theory is a firmly rooted theory with many well-known and revered results, but it is not perfect. Importantly, it has stringent rationality requirements as was described in section 2.1.6, which is a limiting factor for the theory; despite its enticing pliability in describing human interaction, an application with a doubtful rationality role will be more difficult to describe in classical game theoretic terms.

Nevertheless, it is a robust framework to build upon, which we will see in chapter 3. Furthermore, as we will see in chapter 4, it has an intuitive framework for logical analysis. The issues that do arise in classical game theory, therefore, are responsible for some challenges in taking steps towards a logic of evolutionary game theory.

Chapter 3

Evolutionary Game Theory

Classical game theory is centred on the idea that the agents or players of a game reason intelligently and rationally. Humans, being the the exceptionally gifted species on earth capable (we assume) of reasoning intelligently and rationally, are ideal agents around which to build classical game theory. However, there is a manner in which we can include the less "gifted" members of the animal kingdom in game theory. By means of *evolutionary* game theory. This involves considering games where the players are members of a population; the population is divided by the amount n of strategies occurring in the population. Evolutionary game theory measures the circumstances under which the frequencies of a strategy in a populations are stable, and if they are not stable then it measures the dynamic behaviour of the strategies in relation to the other strategies in the population.

Like classical game theory, evolutionary game theory describes concepts by means of the components we saw in the previous chapter: players, strategies, outcomes and preference. The *pivotal* difference separating the concepts of evolutionary game theory apart from those in classical game theory is their interpretation. This difference is the foundation of the arguments in this thesis. In the following chapter 4, we will see how classical game theory has been described with modal logics, but introducing evolutionary game theory to logic is subject to a deep understanding of the consequences of the difference in interpretation. Simply "throwing" the logic we have for classical game theory at evolutionary game theory is inadequate to properly define evolutionary concepts.

This section will thus thoroughly describe how the terminology we are now familiar with is reinterpreted under evolutionary game theory. Following this, the two central approaches to evolutionary game theory are introduced. First, the notion evolutionary stable strategy describes how a homogeneous population of σ–players is robust against any invading mutants. This entails that a mutant, say a τ–player, cannot win in games against the incumbent, the σ–player, at a rate by which the strategy τ "takes over" the population. When a population wards off invasion, it is considered to be evolutionary stable.

The other approach, *replicator dynamics*, describes the dynamic scenario that arises when a population is not robust against invasion and therefore is not or cannot remain homogeneous. The replicator dynamics are based on the differential equations approach to dynamical systems [23]. These dynamical systems represent populations that are in flux; with strategies that sometimes never reach equilibrium and therefore are always approaching or orbiting equilibrium. Such a heterogeneous environment is, by definition, constantly changing over time. For our purposes a brief study of basic calculus is sufficient to describe the rate of change of the frequency over time of a strategy in a population.

Replicator dynamics reflect the intrinsic dynamic factors of evolutionary games, but this thesis' main focus is the static concept of evolutionary stability. The static notion of evolutionary stable strategy "is particularly useful because it says something about the dynamic properties of a system without being committed to any particular dynamic model" [23]. In fact, it is more committed to the static model of a strategic form game; the evolutionary stable strategy is defined similarly to other solution concepts in strategic form games and it can be described by a game matrix. The reason this thesis focuses on strategic form games and the static evolutionary stable strategy is because the logic for strategic form games is the most salient initial doorway to logic for an evolutionary concept.

The main terminology and results in evolutionary game theory that we will see in this chapter is informed by or based on the insightful and thorough sources [42], [28], [23], and [27]. Moreover, altough this chapter thoroughly presents evolutionary stable strategies,s a rigorous discussion of replicator dynamics and dynamical systems is excluded for they are far more involved and complex than is germane to the conclusions of this thesis. For a more in-depth understanding of these topics, [23], [27], and [43] are recommended.

3.1 Background

Evolutionary game theory essentially has its origins in Charles Darwin's theory of natural selection [18], which claims that the fittest genetic (or behavioral) traits are more likely to survive through replication. *Survival of the fittest* easily reflects the game theoretic idea that the interaction[1] of players with certain traits (this is often strictly competitive in natural biology, for animals often compete for limited resources) results in winners and losers. Therefore, the involvement of game theoretic concepts in the process of evolution was a salient direction of study.

The introduction of the paper "The Logic of Animal Conflict" by John Maynard Smith and George Price [42] in 1973 significantly progress in the field of evolutionary game theory. This paper introduced the idea and definition of evo-

[1] After all, "the fitness of an individual cant be measure in isolation; rather it has to be evaluated in the context of the full population in which it lives"[19].

lutionary stability which described why one behaviour rather than another was adopted by a certain proportion of a population at a given time; an idea based on the basic concepts of evolution. Following this, the field became popular among researchers in fields such as economics, language and mathematics, because its new dynamic perspective saliently describe trends in economics, culture, and language.

3.1.1 Evolutionary Interpretation of Game Theoretic Terms

Evolutionary game theory is based on a reinterpretation of the classical game theoretic terminology. In classical game theory, a game Γ is a triple $\Gamma = \langle N, (A_i)_{i \in N}, (u_i)_{i \in N} \rangle$ where N is the set of players, $(A_i)_{i \in N}$ is the set of actions available to i (we defer to the more general notion of strategy as described in Chapter 2), and $(u_i)_{i \in N}$ is the utility function for i, $u_i : A \to \mathbb{R}$. evolutionary game theory reinterprets each of these components in the following way:

> N
> The set N is the set of players that compose a population.
>
> $(A_i)_{i \in N}$
> Instead of choosing an action in a game, players are programmed for certain strategies. This can be interpreted as the expression of a gene or an instinct; namely "parents pass on their strategy [asexually] to their offspring basically unchanged" [28]. This is called *replication*. When two players play a game, each "plays" or enacts its programmed strategy.
>
> $(u_i)_{i \in N}$
> Utility is interpreted as *fitness*. A player who play a game and wins has a higher fitness than his opponent and vice versa. If they "tie" they have the same fitness. A player with higher tness produces more offspring (with the same strategy) than the opponent who has a lower fitness and therefore produces less offspring.

Given this interpretation of game theoretical terms, the following facts and assumptions also hold:

- If a population is heterogeneous with respect to strategies played, we reason about the proportions of each strategy in the population where the sum of the proportion of each strategy is 1.
- Sometimes replication is unfaithful, which results in a mutant who plays a different strategy than his parents. Depending on the utility of his strategy as opposed to the strategy of the incumbent, he will either infiltrate the population or die off.
- Members of the population are paired randomly.
- Birth and death rate are constant.

The way in which the classical game theoretic terms have been interpreted allows the game theorist to define evolutionary notions elegantly. This is evident in the following section, which characterises one of the central notions of evolutionary game theory, the evolutionary stable strategy.

3.2 Evolutionary Stability

Whereas the Nash Equilibrium was the principal solution concept in classical game theory, the evolutionary stable strategy is the principal solution concept of evolutionary game theory. Nash equilibrium expresses that no player has an incentive to to choose a different strategy than the current one. An evolutionary stable strategy (ESS), on the other hand, is a strategy that persists in a homogeneous[2] population and cannot be "invaded" by a group of mutants who play a different strategy. Those mutants "will eventually die off over mutliple generations" [19].

Given the radical reinterpretation of classical game theoretic concepts, Nash equilibrium and ESS describe seemingly separate concepts, but they are, in fact, closely related. An ESS concerns a population of players who each encounter other players in that population and create constantly occurring scenarios of standard 2-player strategic games. Although ESS speaks to the strategies played in a population of players, its definition is entirely based on the utilities for each individual player in simple pairwise, one-shot strategic form games. The definition of an ESS is as follows:

Definition 3.2.1. *A strategy, $\sigma^* \in S$, is an* ***evolutionary stable strategy*** *if for all $\tau \neq \sigma$,*

1. *$u(\sigma, \sigma) \geq u(\tau, \sigma)$, and*
2. *If $u(\sigma, \sigma) = u(\tau, \sigma)$, then $u(\sigma, \tau) > u(\tau, \tau)$*

This definition[3,4] expresses that in order for a strategy σ to be an ESS, it must satisfy two conditions. The first condition states that it must be a Nash equilibrium. That is, the utility of playing σ against σ is better than or equal to playing τ against σ. The second condition expresses that "if a τ-mutation can survive in an σ-population, σ must be able to successfully invade any τ-population..." [28]. That is, if the utility for playing σ against σ is the same for playing τ against σ, then the utility of playing σ against τ must be higher than playing τ against τ. This ensures that when encountering a mutant playing τ, σ will prevail, making it impossible for τ to infiltrate the population of σ players.

[2]With respect to what strategy the members of the population play

[3]Recall that in a symmetric game, $u(A, B) = u_i(A, B) = u_j(B, A)$ for $i, j \in N$; i.e. the utility of strategy profile S for all players in the game.

[4]The following is a trivial variation of the definition of ESS.

1. $u(\sigma, \sigma) > u(\tau, \sigma)$, or
2. If $u(\sigma, \sigma) = u(\tau, \sigma)$, then $u(\sigma, \tau) > u(\tau, \tau)$

3.2.1 A Note on the Relationship Between ESS, NE and SNE

The evolutionary stable strategy is related to Nash equilibrium in a number of ways. Technically, an ESS is also a refinement of a NE; the first condition requires that the ESS is a NE, so naturally that implies that every ESS in a game is also a NE. Thus, ESS $\subset$ NE. Moreover, if a NE is a *strict* NE (SNE), then the requirement for ESS is automatically satisfied. Therefore, we can conclude that [28]:

$$\text{Strict NE} \subset \text{ESS} \subset \text{NE}$$

It is crucial to point out that this claim is superficial for it boldly claims that ESS is a subset of a NE and a superset of a SNE despite the fact that NE and SNE are *strategy profiles* and an ESS is a *strategy*. One must already assume that because ESS holds only for symmetric games, that implies a strategy profile where each player is playing the ESS (say (σ, σ)). This is also astutely noted in [17]:

> ... an ESS must correspond to a symmetric Nash equilibrium in that game ... I say 'correspond' rather than 'is' because an ESS is defined as a single strategy, with the understanding that it is played by all members of a monomorphic population, but a Nash equilibrium is defined as a combination of strategies, one for each player.

Given that the strategy profile corresponding to an ESS must therefore always be symmetric, the above claim fails to specify that not all strict Nash equilibria can be ESSs, but only all symmetric strict Nash equilibria are ESSs. The following game demonstrates that it is possible for a symmetric game to have a strict Nash equilibrium that is not symmetric.

	A	B
A	0,0	3,2
B	2,3	0,0

We must specify that an ESS is, in fact, the superset of a *symmetric* SNE and the subset of a *symmetric* NE. It is also pointed out in [17] that "it can also be shown ... that in this model any strict, symmetric Nash equilibrium corresponds to an ESS in a large population." Therefore, the following is more accurate than the claim above:

$$\text{Symmetric Strict NE} \subset \text{ESS} \subset \text{NE}$$

In any case, we can not simply claim that every SNE is an ESS[5]. Under most circumstances the fact that NE and SNE are strategy profiles, whereas ESS is a single strategy, is relatively trivial but in a logic of evolutionary game theory, it will be crucial to acknowledge the distinction.

[5]We saw that Nash proved that every game has a Nash equilibrium: if not pure, then mixed. [20] claim that a "special case" of Nash's existence theorem is "that every finite [symmeric] game has a 'symmetric' equilibrium."

An evolutionary stable strategy is a special concept in evolutionary game theory. It is the one static concept on which the real dynamic character of evolutionary game theory is built. The following section describes replicator dynamics, which demonstrates the behaviour of strategies when they are not at a stable state. The dynamics are represented by coordinate systems that visualize the changing frequency of strategies over time.

3.3 The Replicator Dynamics

Strategies are considered to be *replicators* and therefore, the dynamics of strategy frequency can be measured over time. If a population has an evolutionary stable strategy, but is for some reason at a heterogeneous state, the replication of strategies over multiple generations will always eventually result in a homogeneous population. However, a population at a homogeneous or heterogeneous state that does not have an ESS also changes over many generations of replication. This results in a scenario where the population never stabilizes and remains in constant flux. In both cases, elementary calculus is sufficient to measure the changing frequency of strategies. This section will introduce the basic mathematics and results that compose the replicator dynamics.

The function $\mathcal{N}(t)$ expresses the population size at a time t. $\mathcal{N}(t)$ "can be thought of as the actual (discrete) population size divided by some normalization constant for the borderline case where both approach infinity" [28]. Given that there are n strategies $\sigma_1, ..., \sigma_n$, $x_i = \mathcal{N}_i/\mathcal{N}$ of some strategy σ_i. The sum of x_j for all $\sigma_j \in S$ is 1; i.e. it is a probability distribution. We also assume that death is a constant d.

Suppose t is continuous. Assuming Δt goes to 0, the limit equation is:

$$\frac{d\mathcal{N}_i}{dt} = \mathcal{N}_i(\sum_{j=1}^{n} x_j u(i,j) - d)$$

"We abbreviate the expected utility of strategy σ_i, $\sum_{j=1}^{n} x_j(i,j)$, as $\overline{u}_i$, and the population average of the expected utility $\sum_{i=1}^{n} x_i \overline{u}_i$, as $\overline{u}$" [28]. Therefore, the corresponding differential equations are:

$$\frac{d\mathcal{N}_i}{dt} = \mathcal{N}_i(\overline{u}_i - d)$$

$$\frac{d\mathcal{N}}{dt} = \mathcal{N}(\overline{u} - d)$$

Because $x_i = \mathcal{N}/\mathcal{N}_i$ (and following from the definition rule for the first derivative), the equation called the *replicator dynamics*:

$$\frac{dx_i}{dt} = x_i(\overline{u}_i - \overline{u})$$

If the differential utility $\frac{dx_i}{dt} = 0$ for all i, the frequency of the strategy is question σ remains constant over time. On the other hand, if it does not equal 0, the strategy will either increase or decrease in frequency. When it is 0, however, the strategy is considered to be static, which reflects the first condition of ESS. The second condition, that the strategy is robust against mutations or invasions of other strategies holds if the trajectories are drawn to the x value resulting in the differential utility of 0. Then x is called an *attractor*.

3.3.1 Graphing Replicator Dynamics

We can visualise the replicator dynamics of a game in a coordinate system where time is mapped to the x–axis and the relative frequency of the strategy in question is mapped to the y–axis. This can be seen in 3.1. figure

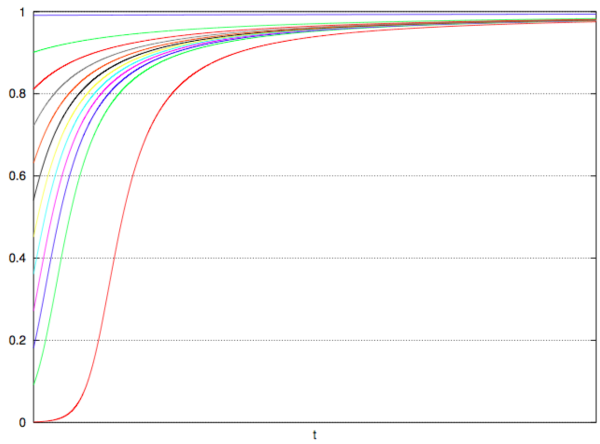


Figure 3.1: This is an example from [28] of a graph representing the replicator dynamics of a game where the frequency of 1 is clearly an attractor for the strategy, meaning that it is an ESS. Each line stands for its behaviour in its trajectory towards a frequency of 1 given some examples of the frequency at which it may "start."

A second way we can depict replicator dynamics is by means of orbital trajectories. In a graph exhibiting orbits, each axis represents the frequency of a strategy and the graph itself has a circular path which depicts a constantly evolving, never stabilizing, replicator dynamics. There is no attractor, therefore, there is no ESS.

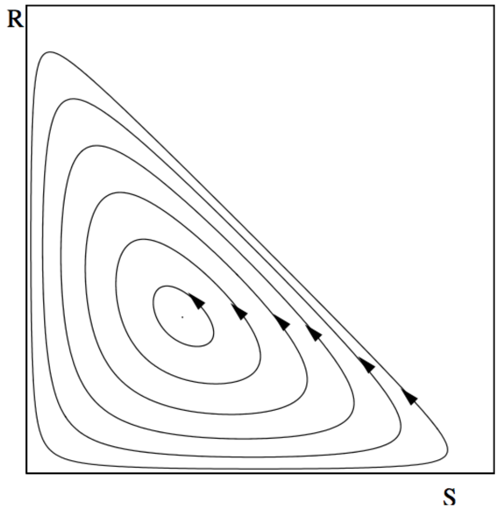


Figure 3.2: This example, also borrowed from [28], is a game of three strategies, where the frequency of the strategies are constantly evolving. At some point on a trajectory, the next generation has on average evolved to play less of the previous generation's strategy.

3.4 Conclusion

This chapter introduced evolutionary game theory with an intentional focus on its relationship with classical game theory. Classical game theory provides the framework and machinery for evolutionary game theory. We saw in this chapter that players, strategies, strategy profiles and utilities formally operate in the same way.

The main difference is that evolutionary game theory takes this machinery a step further by reinterpreting the meaning of the terms and the machinery it has adopted from classical game theory. In particular, because it measures utility as fitness, evolutionary game theory inherits a future-oriented picture of game theory. Fitness implies something about how the same game may be played in future encounters between members of a population, where the population changes after every occurrence of the game.

So although evolutionary game theory is actually very (structurally) similar to classical game theory, but also very different (in meaning). This difference in meaning more consequential for its relationship with classical game theory; that is, this difference will manifest itself in building a logic for evolutionary game theory out of logics for classical game theory. In the next chapter, we will see some existing logics for classical game theory. My selection of sources in the next chapter will be conducive to the upcoming proposals for an evolutionary game logic. This will show that classical game logics establish a reliable core for evolutionary game logics, but the change in meaning demands the same from an appropriate logic.

This chapter also introduced the notion of dynamical systems and replicator dynamics in evolutionary game theory. Although replicator dynamics demonstrate a central point of evolutionary game theory, i.e. change in strategy frequency over time, it will not be a widely discussed factor in this thesis. This is certainly a topic of future work for logicians, where a logical study of dynamical systems would require a way to describe infinite games [6], Chapter 5 for an overview), and a way to express continuous functions. Kremer and Mints [30] do suggest a logic to this end, in terms of "dynamic topological logic." They base the logic on topological spaces in place of Kripke frames.

The primary directive of this thesis, however, is on the more static concepts in evolutionary game theory such as evolutionary stable strategy. ESS raises many interesting questions as a static concept alluding to a dynamic environment. The fact that we can (and will) define ESS in modal logic (which seems extraordinarily simple compared to logics that express the replicator dynamics). Nevertheless, logical applications alluding to the temporal nature of replicator dynamics is not ignored.

[6] See [4]

Chapter 4

Logic in Classical Game Theory

Logic and classical game theory have many things in common; they both aim to model complex interactive scenarios composed of concerns rooted in both classical game theory and logic, such as agents, choices, actions, preference, and knowledge (or lack thereof). For some time now, logicians have been exploring the enticing possibilities as well as the challenges posed within the interface of classical game theory and logic.

Van Benthem's distinction between logic games and game logics [4] outlined in the introduction draws a line between topics within the interface. On one side, logic games consider classical game theory as a tool used to describe logic, where, in short, players are burdened to "prove" the truth of logical formulas. The player with the winning strategy determines whether a formula is true or false. Jaakko Hintikka and Gabriel Sandu have published a seminal and elaborate work, "Game Theoretic Semantics" [25], on this topic.

On the other hand, game logics aim to model a game as a framework of rationality and interaction by means of logic. This section will take this latter perspective by analysing a selection of approaches to modelling classical game theory in both strategic and extensive forms. In particular, this chapter focuses on modal logic with epistemic modalities; one static modal logic approach and one dynamic modal logic approach. We see that the procedure of IEDS is an ideal illustration of the kind of rational interaction that modal logic can describe.

The interface between logic and classical game theory is still a rapidly growing field, for logicians are inventing ever more creative ways to describe the workings of classical games. Therefore, this chapter will also investigate Jeremy Seligman's novel approach [40] to describing the structure of (strategic) games by means of an extension of modal logic, *hybrid logic*. This approach, as well as a recent development on this approach by van Benthem, will compose a significant portion of this chapter.

In accordance with the trend of creative approaches to logic and games, this chapter will also build on Seligman's [40] and van Benthem's [10] [4] work by demonstrating how hybrid logic can describe many more interesting concepts in classical game theory such as the salient example, IEDS.

In the next chapter, the techniques investigated in this chapter will see yet another advance under *evolutionary* game theory. Evolutionary game theory is not as intuitive to logic as classical game theory is, but this thesis nevertheless takes the hybrid logic development described here to be an attractive way to introduce logic to evolutionary game theory.

4.1 Logic for Strategic Form Games

This section will explore several contributions that offer compelling logics for strategic games. One particular challenge for logicians arises from the exclusion of epistemic factors in the traditional definitions of strategic games. Nash equilibrium, best response and the process of iterated elimination of dominated strategies (IEDS)[1], for instance, have stringent requirements for the knowledge and rationality of the players. In order for them to make sense realistically, one must assume common knowledge of rationality for the players. Often game theorists, at best, give informal arguments for the epistemic foundations of these solution concepts. For instance, Osbourne and Rubenstein, in their influential introduction to game theory, describe that a player "is aware of his alternatives, forms expectations about any unknowns, has clear preferences and chooses his action deliberately after some process of optimization" [35]. Others have made similar descriptions:

> Game theory has originally been conceived as a theory of strategic interaction among fully rational agents ... Rationality here means, among other things, full awareness of ones own beliefs and preferences and logical omniscience. Even stronger, for classical game theory to be applicable, every agent has to ascribe full rationality to each other agent. [29]

The articles surveyed in this section are Giacomo Bonanno's "Modal Logic and Game Theory: Two Approaches"[13], van Benthem's "Rational Dynamics and Epistemic Logic in Games" [2], van Benthem et al.'s "Towards a Theory of Play" [10] and Seligman's "Hybrid Logic for Analysing Games" [40], describe strategic games using modal logic. Each article does this in a unique way. With this selection, we describe strategic games in static logic, strategic games in dynamic logic and strategic games in a hybrid logic.

In section 6.1.1, I outline how Giacomo Bonanno describes games in static modal

[1] See Chapter 2.

logic; in particular by his modal logic perspective of IEDS. Van Benthem et al. address IEDS as well, by means of dynamic modal logic. This will be discussed in section 6.1.2. These sources both interpret important epistemic characteristics about strategic games through modal logic. These characteristics are bundled together into a new relational structure which is very expressive, and in section 6.1.3, we will describe these epistemic characteristics and formalize them with the help of [10] and [40]. Last, in section 6.1.4, I will introduce hybrid logic, which I consequently use to describe many of the concepts already discussed in this chapter. Hybrid logic, which "brings to modal logic the classical concepts of identity and reference" [12], is an elegant yet powerful way to describe strategic games. It proves to express game theoretic concepts from best response to IEDS to symmetric games.

4.1.1 Games in Static Modal Logic

Giacomo Bonanno provides an early example [13] of a modal logic for strategic games. He bases the logic on a Kripke frame $\langle \Omega, R_1, ..., R_n, R_* \rangle$ where Ω is a set of states, R is a binary relation on Ω, $\{1, ..., n\}$ is the set of players and the $(n+1)th$ relation R_* is the transitive closure of $R_1 \cup ... \cup R_n$. However, the interpretation of the relation R depends on the interpretation of game theory he takes. One interpretation views "game theory as a description of how rational individuals behave" and the other views it as "a prescription ... to players on how to act."

Game theory as a description of rational behaviour seeks a way to account for knowledge (of the game and other players) and rationality. Because the standard view of games (in this case, finite non-cooperative strategic form games) "provides only a partial description of the interactive situation" [13], where beliefs and rationality are not addressed, Bonanno devises a system where they are. In order to "illustrate the types of results obtained" [13] from his formalism, he uses it to describe how a strategy profile survives IEDS in it.

Bonanno adds a probabilistic element to the above frame to account for players' belief. We get the game model: $\langle \Omega, R_1, ..., R_n, R_*, P_1, ..., P_n \rangle$ where Ω and $\{1, ..., n\}$ are as above. In this interpretation, we specify R to be associated with the modal formula $\Box_i A$ which denotes "player i believes that A" and is true at a state $\alpha \in \Omega$ iff A is true at every state β such that $\alpha R_i \beta$. The intended interpretation of R is "for player i state β is epistemically accessible from state α " [13]. R_* is associated with $\Box_* A$ which denotes "it is common belief that A." P_i is the probability distribution on Ω and is used to describe belief; $p_{i,\alpha}$ which denotes player i's belief at state α is defined by conditioning P_i on $R_i(\alpha) = \{\omega \in \Omega : \alpha R_i \omega\}$:

Definition 4.1.1. *Player i's **belief** at state α for the worlds R_i-accessible from α:*

$$p_{i,\alpha}(\omega) = \begin{cases} \frac{P_i(\omega)}{\sum_{w' \in R_i(\alpha)} P_i(\omega')} & \text{if } \omega \in R_i(\alpha) \\ 0 & \text{if } \omega \notin R_i(\alpha) \end{cases}$$

In other words, if ω is in the set of states R-accessible from α, then the probability (belief) of ω for i is a fraction of the total amount of states accessible from α.

The model also includes a valuation represented by the pair (V, σ) where V "associates with every atomic proposition the set of states at which the proposition is true" and σ is a function $\sigma = (\sigma_1, ..., \sigma_n) : \Omega \to S$ "that associates with every state the *pure* strategy profile played at that state" [13]. There are two integral atomic propositions that refer to facts about strategy profiles: r_i expresses that i is rational, and s^∞ is a strategy profile in the set of strategy profiles S^∞ that survive the process of IEDS. Suppose the following restrictions hold for (V, σ):

1. (a) $\alpha \in V(r_i)$ iff $\alpha R_i \beta \Rightarrow \sigma_i(\beta) = \sigma_i(\alpha)$
   (Player i has no uncertainty of the strategy he himself plays.)
   (b) $\sigma_i(\alpha)$ maximizes i's expected utility given his beliefs.
2. $\alpha \in V(s^\infty)$ iff $\sigma(\alpha) \in S^\infty$
   (The strategy played at α survives IEDS.)

It is then possible to introduce the formula $\Box_* r \to s^\infty$, which claims that "if there is common belief that all the players are rational, then the strategy profile actually played is one that survives the [IEDS]" [13].

Now we turn to Bonanno's second interpretation of game theory: as a normative model for which solutions are recommendations to players on how to act. In this interpretation, the meaning of the relation R changes. Here, $\alpha R_i \beta$ denotes "from state α player i can unilaterally bring about state β. Thus R_i does not capture the reasoning or epistemic state of player i but rather the notion of what player i is *able to do*" [13]. R_* is interpreted as "at state α it is recommended that state β be reached."

In this interpretation, there are also new atomic propositions, where $p, q \in \mathbb{Q}$); $(u_i = p_i)$ expresses "player i's utility is p_i," $(q \leq p)$ expresses "q is less than or equal to p," and the proposition *Nash* expresses "the pure strategy profile played is a Nash equilibrium." Last, the pair (V, σ) satisfy a new set of restrictions:

1. $\alpha R_i \beta$ iff $\sigma_{-i}(\beta) = \sigma_{-i}(\alpha)$
2. If a is an atomic proposition of the form $(q \leq p)$, then $V(a) = \Omega$ if $q \leq p$ and $V(a) = \emptyset$ otherwise.
3. $\alpha \in V(u_i = p_i)$ iff $u_i(\sigma(\alpha)) = p_i$
4. $\alpha \in V(\textit{Nash})$ iff $\sigma(\alpha)$ is a Nash equilibrium of the game.

The modal formulas $\Box_i A$ and $\Box_* A$ get new interpretations as well. Those are "no matter what unilateral action player i takes, A is true" and "it is recommended that A," respectively. The recommendation to play a Nash equilibrium is defined as:

$$\Box_*(\bigwedge(u_i = p_i) \rightarrow \bigwedge \Box_i((u_i = q_i) \rightarrow (q_i \leq p_i)) \rightarrow \Box_*(Nash)$$

It is generally agreed upon that the strategy profiles in a game are represented as states in a Kripke frame. Bonanno claims that "in order to obtain a model of a particular game G, [...] we also need to add a function that associates with every state the *pure* strategy profile played at that state" [13].

4.1.2 Games in Dynamic Modal Logic

As Bonanno did in his first interpretation of game theory, van Benthem [2] also recognizes that IEDS is an interesting epistemic process that can be depicted in logic. However, he uses a more modern logic, dynamic epistemic logic, to illustrate the dynamic process of IEDS and the epistemic components of game theory it reveals. Van Benthem's motivation is that game solution algorithms such as IEDS can be seen as a process of model update based on what the agents know and learn about each other. Therefore dynamic epistemic logic is an appropriate tool to describe this.

It is interesting to note that Bonanno's analysis of IEDS actually also has a dynamic flavor. Above, S^∞ was described as the set of strategy profiles that survive the process of IEDS. In fact, S^∞ is the set of strategy profiles in G^∞, or "the game obtained by applying [the] iterative elimination procedure" [13]. That is, given an initial game G^0, G^1 is obtained by deleting the pure strategies strictly dominated in G^0. Continuing in this way, G^n for $n \geq 1$ is obtained by deleting the pure strategies strictly dominated in G^{n-1} until reaching G^∞, which is the game that results when no more pure strategies can be deleted. Thus G^∞ is a subgame of G^0 obtained by a process that is essentially a model update prompted by the application of a deletion. Van Benthem continues along this line of reasoning by characterizing exactly what prompts a move of deleting strictly dominated strategies. Using basic epistemic logic and public announcement logic, he outlines a system where an announcement of rationality "triggers" a model update.

Basic epistemic logic is a propositional logic together with modal operators $K_i\varphi$ denoting "i knows φ" and $C_G\varphi$ denoting "φ is common knowledge in group G." We have the standard picture of reasoning from within a model $\mathcal{M}$ and in a current world w such that $\mathcal{M}, w \models \varphi$ denotes that in model $\mathcal{M}$, φ is true at the current world w. The following hold for the modal operators:

$\mathcal{M}, w \models K_i\varphi$ if and only if for all v with $w \sim_i v$, $\mathcal{M}, v \models \varphi$

$\mathcal{M}, w \models C_G\varphi$ if and only if for all v that are accessible from w by some finite sequence $\sim_i$ steps (any $i \in G$), $\mathcal{M}, v \models \varphi$

Where the relation $\sim_i$ expresses an agent i's epistemic accessibility.

The dynamic element of the game logic is based on public announcement logic which is a logic that describes model change after the occurrence of an action, communication. $[P!]\varphi$ is the dynamic modality that expresses "after a truthful public announcement of P, formula φ holds" [2], with as truth condition:

$$\mathcal{M}, w \models [P!]\varphi \text{ if and only if, if } \mathcal{M}, w \models P \text{ then } \mathcal{M}|P, w \models \varphi$$

Together with the epistemic logic we can write formulas such as $[A!]K_i\varphi$ expressing that "after an announcement of A, i knows that φ," and $[B!]C_G\varphi$ expressing that "after an announcement of B, φ is common knowledge for group G."

To build the language of a game logic, van Benthem thus adjusts the standard picture of a strategic game, $G = \langle N, (A_i)_{i\in N}, (\precsim_i)_{i\in N}\rangle$ to a model over G; that is, $\mathcal{M}(G)$, where strategy profiles are worlds, and $\sim_i$ is the epistemic accessibility relation that i has over some worlds is the game. A player i can only have this relation over worlds that he can tell apart. Those worlds are the ones where i's opponents' strategies stay the same, but i's change. After all, i has the ability to differentiate and consequently choose between the worlds in this relation by simply changing his own strategy. For example, for strategy profiles $\sigma = (a_i, a_{-i})$ and $\sigma' = (a_i, a'_{-i})$, it holds that $\sigma \sim_i \sigma'$.

To describe solutions such as best response and Nash equilibrium, there are additional terms in the language: i plays action $\omega(i)$ in world ω, and $\omega(i/a)$ denotes the strategy profile ω where i replaces action $\omega(i)$ with a.

Best response is expressed as a proposition B_i saying that "i's utility cannot improve by changing her action in ω – keeping the others' actions fixed" [2].

Definition 4.1.2. ***Best response*** *for i is true in ω if*

$$\mathcal{M}, \omega \models B_i \text{ iff } \bigwedge_{a\in A_i|a\neq\omega(i)} \omega(i/a) \precsim_i \omega$$

In other words, all instances where i plays some a instead of $\omega(i)$ is less preferred by i. Consequently, Nash Equilibrium is expressed as the conjunction of best responses.

Definition 4.1.3. ***Nash Equilibrium*** *is true in ω if*

$$\mathcal{M}, \omega \models NE \text{ iff } \bigwedge B_i \text{ for all } i \in N.$$

These standard definitions of best response and Nash equilibrium must be adjusted in order to account for model changes after updates. If we define best response based on worlds that are (left in) a model after an update, instead of the full game model with which we started, we get a dynamic notion for best response as well:

Definition 4.1.4. *The proposition B_i^*,* ***relative best response*** *for i, is true at a world if it is a best response to $-i$ where the range of alternative strategies is limited to all other strategies in $\mathcal{M}$.*

Relative best response therefore may change when the model changes.

In games, rationality plays as big of a role as knowledge does, for together they justify the solution concepts. An agent is considered rational if he, in fact, plays his "best response given what [he] knows or believes" [2]. van Benthem demonstrates that we can relativize rationality with regards to model updates as well.

The relative version of rationality, *weak rationality*, is based on what the agent knows about the game, and may therefore be limited to a smaller model than the original full game.

Definition 4.1.5. ***Weak Rationality.***
$\mathcal{M}, \omega \models \mathrm{WR}_i$ *iff* $\bigwedge_{a \neq \omega(i)} \langle i \rangle$ *'i's current action is at least as good for i as a.'*

The only worlds where this assertion is false is at worlds that are strictly dominated for i. On the other hand, the variant of rationality that is based on the full game model is *strict rationality.*

Definition 4.1.6. ***Strong Rationality.***
$\mathcal{M}, \omega \models \mathrm{SR}_i$ *iff* $\langle i \rangle \bigwedge_{a \neq \omega(i)}$ *'i's current action is at least as good for i as a.'*

Weak rationality is sufficient to characterizing IEDS. By applying public announcement, we see that a repeated alternating announcement of weak rationality by the players will result in an updated model containing exactly the set of worlds that survive the process of IEDS. The theorem is as follows:

> *The following are equivalent for worlds ω in full game models $\mathcal{M}(G)$:*
>
> 1. *World ω is in the IEDS solution zone of $\mathcal{M}(G)$.*
> 2. *Repeated successive announcement of Weak Rationality for players stabilizes at a submodel $\mathcal{N}(G)$ whose domain is that solution zone.*

The repeated announcement of strong rationality, on the other hand, results in a "new game-theoretic solution procedure, whose behaviour can differ" [2]. It eliminates actions that are never best responses from the model. Nash equilibria but also other states survive this process.

4.1.3 Relations in Strategic Games

Modal logics such as those described above demonstrate but do not explicitly identify a salient feature of games that the standard view of classical game theory does not: *relations between the outcomes of a game.* This section will identify the three essential relations that efficiently describe epistemic and rational aspects of strategic games.

The Knowledge Relation

Consider the relationship between two outcomes where i's strategies are the same in both, but $-i$'s strategies are not the same. Player i cannot distinguish between these outcomes, because although he knows his own strategy, he does not know which strategy $-i$ will play. This relation is what we will refer to as *knowledge*, expressing which outcomes are epistemically accessible[2] to i.

Standard game theory generally lacks the language to formally express facts about the players' knowledge of a game. We have seen that modal logic, on the other hand, does already have the machinery to reason about it by taking knowledge as a relation between outcomes. But how do we define this relation to characterize a player's knowledge? First we can refer to the above surveys of [13] and [2], where both Bonanno and van Benthem already formally express the the knowledge relation in their logics.

Bonanno needs this relation to ensure the correct interpretation of the rationality atom r. He postulates that a strategy profile α is a state where r_i is true if and only if $\alpha R \beta$ implies that player i knows what strategy he is playing. In other words, $\sigma_i(\alpha) = \sigma_i(\beta)$ (i plays the same strategy in strategy profiles α and β). Player i is thus rational if these truth conditions hold *and* player i plays a strategy in α that maximizes his expected utility given his beliefs. As we have seen, Bonanno needs this to characterize IEDS.

Van Benthem also identifies the knowledge relation between worlds where "players know their own action, but not that of the others" [2]. In this case, knowledge, encoded by symbol $\sim_i$, is also integral to characterizing IEDS. Van Benthem introduces some interesting properties about the knowledge relation. It establishes an epistemic foundation for the grid structure of a matrix game by composing the relations of each player. The knowledge relation runs over rows for the row player, and over the columns for the column player.

The Confluence Axiom [2] $\sim_i \sim_j \varphi \rightarrow \sim_j \sim_i \varphi$ indicates that any world of the game can be reached; the composition of$\sim_i$ and $\sim_j$ is therefore universal accessibility.

The Preference Relation

Bonanno had, as a second condition for the proper interpretation of rationality, that player i plays a strategy in α with maximal utility. Van Benthem also brings utility into the picture with regard to (relative) best response and rationality. Recall that best response B_i is true in a world if, given his opponent j's strategy, i cannot improve his utility by changing his own strategy. The WR loops are based on the idea that a world with a best response for i is $\sim_j$ accessible to a world with a best response for j which is $\sim_i$ accessible to a world with a best response for i ... etc. More specifically, "given a world with some action for i, there must be a world in the model with that same action for i where j's utility is highest" [2]; in other words, $\langle i \rangle B_j^*$.[3]

[2]To prevent confusion, although this relation is called knowledge, it actually the epistemic relation that indicated that a player cannot distinguish between two states. He only *knows* what strategy he, himself, is playing

[3]If we repeatedly apply this fact to a 2-player game, it will result in a sequence $B_i^* \sim_i B_j^* \sim_j B_i^* \sim_i B_j^* ...$ which will loop, because it is a finite game. "Looking backwards" over

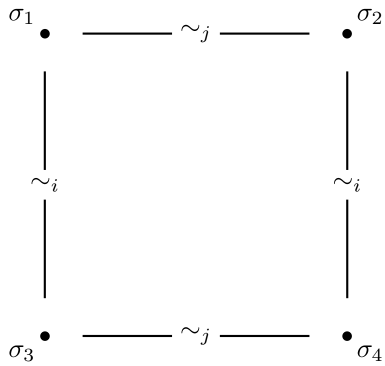


Figure 4.1: This shows that the $\sim_i$ relation runs vertically and the $\sim_j$ runs horizontally.

From these examples we observe that $\sim_i$ is not sufficient to describe features of games and justifiably so. After all, a cornerstone of game theory is preference; a player in a strategic scenario is motivated by the prospect of the best possible outcome. We will therefore consider *preference* to be a relation between outcomes just as with the knowledge relation described above[4]. The preference relation will be expressed by the symbol $\succeq$ where $\sigma \succeq_i \sigma'$ expresses that i prefers σ equally to or more than σ'.

Of course, the notion of preference is not novel. There is already a large body of literature behind preference logic, and essentially we can rely on the basic definitions and fact that arise from that. In chapter 2, preference was defined as part of a game model Γ. That is essentially what I claim is a "preference relation" here.

The Freedom Relation

There is one more relationship that we can claim arises in a game. The relation knowledge for i connects outcomes where i's strategy is unchanged, but $-i$'s strategies differ, but is also possible to label the connection between outcomes where instead i's strategies differs and $-i$'s is constant. This relation is labelled *freedom*, for it indicates that i has the freedom to directly choose between the outcomes by choosing which strategy to play. Formally, $\sigma \approx_i \sigma'$ expresses "i

this loop, we get a WR loop:

$$s_1 \sim_i s_2 \sim_j s_3 \sim_i s_4 \text{... where } s_1 \models B_j^*, s_2 \models B_i^*, s_3 \models B_j^*, s_4 \models B_i^*$$

[4]Note that preference and outcome are related as explained in Chapter 2.

has the freedom to unilaterally bring about σ or σ'".

Bonanno describes this concept with regard to his second interpretation of game theory (prescriptive):

> ... the interpretation that we want to establish for $\alpha R_i \beta$ is no longer "for player i state β is epistemically accessible from ... state α but rather "from state α player i can *unilaterally* bring about state β." Thus R_i does not capture the reasoning or epistemic state of player i but rather the notion of what player i is *able to do.* [13]

On the other hand, in [2] as well as in standard descriptions of game theory, the freedom relation is more implicit. Best response as an atomic proposition true at some state σ, in fact, expresses that there is *no* outcome $\approx_i$ accessible from σ that i prefers more than σ'. Thus best response coincides with freedom's stipulation that $-i$'s strategy is held constant.

Jeremy Seligman had originally pinpointed this relation and named it freedom [40]. In fact, he considered freedom (F) as well as the other relations, knowledge (K) and preference (P), as follows:

> $F(w, u, a, t)$: worlds w and u are identical up to time t, and at that time they are identical but for a free choice of agent a.
>
> $K(w, u, a, t)$: at time t, agent a cannot distinguish worlds w and u.
>
> $P(w, u, a, t)$: at time t agent a regards world u no worse (and possibly better than) world w.

F corresponds to our $\approx$, K to our $\sim$, and P to $\succeq$. In addition to identifying these relations, Seligman reveals some interesting facts about them. For instance, $K(u, v, a, t)$ iff $F(u, v, b, t)$ for some agent b other than a. In others terms, $u \sim_i v \Leftrightarrow u \approx_j v$ for $i \neq j$. Freedom is an interesting new concept which will be one main focus of this thesis; an explicit way to refer to players' strategies is relevant.

Seligman also describes that these relations can be used as modal operators $\langle F_i \rangle$, $\langle K_i \rangle$, and $\langle P_i \rangle$ for each agent $i \in A$ where

$$\mathcal{M}, w \models \langle F_i \rangle \varphi \text{ iff } F(w, v, i) \text{ and } \mathcal{M}, v \models \varphi \text{ for some } v \in W$$

Seligman finds this language too limited to define many concepts of game theory and therefore continues to develop a system using hybrid logic. This will be discussed later on.

One Relational Structure

Van Benthem et al.[10] have developed Seligman's ideas by bringing $\sim$, $\approx$, and $\succeq$ together into this relational structure (where S is the set of strategy profiles):

$$\mathcal{M} = \langle S, \{\sim_i\}_{i \in N}, \{\approx_i\}_{i \in N}, \{\succeq_i\}_{i \in N} \rangle$$

They use the concepts freedom, knowledge and preference as modalities to develop "a full modal logic of strategic games" [10]. They also pay special attention to the interaction of the modalities and additional methods available to define various concepts in games.

The modalities are based on the concepts discussed above. We formally define the relations as follows:

Definition 4.1.7. *Outcomes $\sigma, \sigma' \in S$ express the relation* ***Freedom****,* ***Knowledge*** *or* ***Preference*** *when the following definitions hold:*

- *Freedom.* $\sigma \approx_i \sigma'$ *iff* $\sigma_{-i} = \sigma'_{-i}$
- *Knowledge.* $\sigma \sim_i \sigma'$ *iff* $\sigma_i = \sigma'_i$
- *Preference.* $\sigma \succeq_i \sigma'$ *iff "player i prefers the outcome σ at least as much as outcome σ'."[10]*

The formula φ expresses facts that hold at a particular state σ. States $\sigma_1, \sigma_2, ..., \sigma_n$ are considered to be exactly the outcomes of the game, and $\sigma \models \varphi$ iff φ is true at outcome σ; φ describes a fact that holds at σ. Specifically, φ is built from a set of atomic propositions p_i^a expressing "player i plays action a." Therefore if outcome $\tau = (\tau_i, \tau_{-i})$, then $\tau \models p_i^{\tau_i}$. A semantics for this and the three relations is as follows:

Definition 4.1.8. ***Semantics.***

$$\sigma \models p_i^a \Leftrightarrow \sigma_i = a_i$$

$$\sigma \models \varphi \wedge \psi \Leftrightarrow \sigma \models \varphi \textit{ and } \sigma \models \psi$$

$$\sigma \models \varphi \vee \psi \Leftrightarrow \sigma \models \varphi \textit{ or } \sigma \models \psi$$

$$\sigma \models \neg\varphi \Leftrightarrow \sigma \not\models \varphi$$

$$\sigma \models \varphi \rightarrow \psi \Leftrightarrow \sigma \not\models \varphi \textit{ or } \sigma \models \psi$$

$$\sigma \models [\sim_i]\varphi \textit{ iff for all } \sigma'\textit{, if } \sigma \sim_i \sigma'\textit{, then } \sigma' \models \varphi$$

$$\sigma \models [\approx_i]\varphi \textit{ iff for all } \sigma'\textit{, if } \sigma \approx_i \sigma'\textit{, then } \sigma' \models \varphi$$

$$\sigma \models \langle\succeq_i\rangle\varphi \textit{ iff there exists } \sigma' \textit{ such that } \sigma' \succeq_i \sigma \textit{ and } \sigma' \models \varphi.$$

$$\sigma \models \langle\succ_i\rangle\varphi \textit{ iff there exists } \sigma' \textit{ such that } \sigma' \succeq_i \sigma \textit{ and } \sigma \not\succeq_i \sigma' \textit{ and } \sigma' \models \varphi.$$

Recall the Stag Hunt game (here with the row marked for player i and column marked for player j).

	Stag_j	Hare_j
Stag_i	2, 2	0, 1
Hare_i	1, 0	1, 1

For this game, we have two players i and j, four worlds σ_1, σ_2, σ_3, and σ_4. With the above machinery we can conclude many facts to be true in each world. For instance

$$\sigma_1 \models p_i^{Stag} \wedge p_j^{Stag} \wedge \neg p_i^{Hare} \wedge \neg p_j^{Hare} \wedge (p_i^{Stag} \vee p_i^{Hare} \wedge p_j^{Stag} \vee p_j^{Hare}) \\ \wedge [\sim_i](p_i^{Stag} \wedge (p_j^{Stag} \vee p_j^{Hare})) \wedge \langle \approx_j \rangle p_j^{Hare} \wedge \neg \langle \succ_j \rangle p_i^{Stag}$$

All of the formulas that following from the above formulas are also true in σ_1. Therefore, there are many propositions that hold in each state of the game.

The following graph depicts how these three relations for strategic games connect the states in a game.

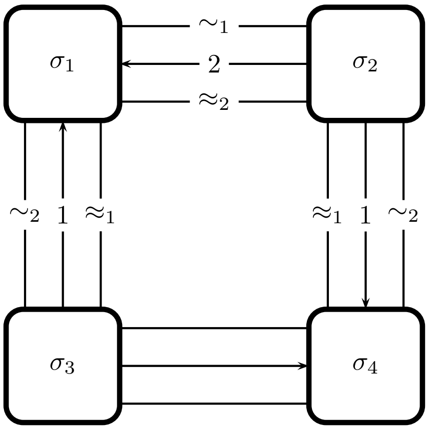


Figure 4.2: This figure depicts four worlds and the relations $\approx$ and $\sim$ for both players as well as some of their preference relations.

Given these nice modalities, van Benthem et al. express the appeal of their interaction. They define the *universal modality*, $[\sim_i][\approx_i]\varphi$ which makes φ true in every world of the model. Furthermore, the following rule, which in essence expresses the "grid property" of game matrices is valid in game models [10]:

Grid Property. $[\sim_i][\approx_i]\varphi \leftrightarrow [\approx_i][\sim_i]\varphi$

In the Stag Hunt, the universal modality makes formulas such as $p_i^{Stag} \vee p_i^{Hare} \wedge p_j^{Stag} \vee p_j^{Hare}$ and tautologies ($\top$) true in every state of the game. Another rule we can add expresses the relationship between a player's *knowledge* and his opponent's *freedom* (or vice versa):

KFeq. $[\sim_i]\varphi \leftrightarrow [\approx_j]\varphi$

In other words, the states over which player i has the knowledge relation, player j has the freedom relation. With this rule, we see that the universal modality

$[\sim_i][\approx_i]\varphi$ is related to the Confluence Axiom that van Benthem described in [2].

In addition to the language so far, we have the tool *intersection modality* [10]. Consider the following formula intersecting freedom and strict preference [10]:

$$\mathcal{M}, \sigma \models \langle \approx_i \cap \succ_i \rangle \varphi \text{ iff for each } \sigma' \text{ if } \sigma(\approx_i \cap \succ_i)\sigma' \text{ then } \mathcal{M}, \sigma' \models \varphi$$

By intersecting the relations we can limit the sets of states to express particular concepts. The above formula intersecting freedom and preference expresses the states that player i's ability to choose the state he prefers. The definition of best response uses this notion [10]:

Definition 4.1.9. *The **best response** for player i is defined as:*

$$\neg \langle \approx_i \cap \succ_i \rangle \top$$

The intersection of $\approx$ and $\succ$ limits the set of accessible strategy profiles to the ones that i can choose between and prefers the most. Negating this modality ensures the correct meaning, "there is no strategy profile better than the current one for i that i can choose."

This relational structure exploiting $\sim$, $\approx$ and $\succ$ gives us a powerful language to formally describe games. In the following section, we will see that using this language together with hybrid logic gives us even more expressive power.

4.1.4 Games in Hybrid Logic

Jeremy Seligman, who demonstrated $\sim$, $\approx$ and $\succeq$ in a strategic game model, concludes that considering them as modalities is too limited to express many game theoretic concepts [40]. He therefore builds a language using elements from *hybrid logic*, which results in a much more expressive and powerful language.

Hybrid Logic

We are already familiar with standard propositional logic and modal logic. Propositional logic is founded on the set of propositions $\mathsf{PROP} = \{p, q, ...\}$ which are formulas on the object level. We also consider the set $\mathsf{MOD} = \{\pi, \pi', ...\}$ where π is some modal operator. This is the standard logical machinery we have been using up to this point.

Hybrid logic, on the other hand, in addition to PROP and MOD, adds a set NOM of *nominals* that allows us to name and refer specifically to some state. NOM is comparable to PROP, but the important difference is that NOM is a non-empty set of variables $a, b, c...$ that are only true at *one state* in a model (whereas members of PROP can be true at many states). A nominal therefore " 'name[s]' this point by being true there and nowhere else" [12].

A hybrid model is a triple $\langle W, \{R_\pi \mid \pi \in \mathsf{MOD}\}, V\rangle$ where each R_π is a binary relation on W, and V is a valuation function $V : \mathsf{PROP} \cup \mathsf{NOM} \to \wp(W)$

such that for all nominals a, $V(a)$ is a singleton subset of W. This valuation function replaces the one we normally use in propositional logic, namely $V : \mathsf{PROP} \to \wp(W)$

Hybrid logic also adds a satisfaction operator, $@_a\varphi$ to the language, which enables us to "jump" to a state in a model named by some nominal a and consequently evaluate any formula with the world named by a as the point of evaluation. The formula $@_a\varphi$ expresses that "φ is true at the state named by a." Formally,

Definition 4.1.10. ***Satisfaction.***

$$\mathcal{M}, w \models @_a\varphi \textit{ iff } \mathcal{M}, w' \models \varphi \textit{ where } w' \in V(a)$$

Naming and referring to particular states in a model is one function of hybrid logic, but we also have the ability to locally name worlds relative to the current point of evaluation. For this, hybrid logic employs binders denoted by a down arrow operator $\downarrow p$ which binds p to the current state (makes p true in the current state). Binders can also be used in combination with relations, where $R \downarrow p$ "binds p to the set of R-successors" [40]. Formally,

Definition 4.1.11. ***Binder.***

$$\mathcal{M}, w \models R \downarrow p\ \varphi \textit{ iff } \mathcal{M}^p_{Rw}, w \models \varphi \textit{ where } Rw = \{v \in W \mid Rwv\}$$

M^p_{Rw} is the submodel of M, restricted to only the R-accessible worlds from w, where proposition p has been "assigned." Assigning a proposition to hold at a set of states in the model does not change the hybrid valuation function V, because it is a matter of insisting that p is true in these states, not observing that they are true. The perspective hybrid logic asks us to take is to think of the model as something being built, not discovered.

"Modal Logic with Binders"

Seligman primarily uses binders in his application of hybrid logic. Binders give us a convenient ability to "hand-pick" worlds according to how they are related to the current world (or the point of evaluation).

If a world w' is accessible from the current world w by multiple relations R, R', R'', w' will satisfy a new proposition with every act of binding, $R \downarrow p$, $R' \downarrow q$, and $R'' \downarrow r$. It follows then that $\mathcal{M}, w' \models p \wedge q \wedge r$.

In this way, we can hand-pick worlds in a model by defining whether they are or are not accessible via certain relations. For example, suppose that xBy indicates that dalmatian x is bigger than dalmatian y, and xSy indicates that dalmatian x is spottier than dalmatian y. If a dog breeder wants a small spotty dalmatian, the dog should satisfy the following formula. Dog x is what the breeder wants if $\mathcal{M}, x \models B \downarrow p\varphi\ S \downarrow q\varphi\ (\neg p \wedge q)$.

Using binders in combination with the global modality E, Seligman demonstrates that it is possible to describe many game theoretic concepts. The global modality $E\varphi$ expresses "there exists a world in the model where φ holds." Formally,

Definition 4.1.12. ***Global Modality.***

$$\mathcal{M}, w \models E\varphi \textit{ iff } \mathcal{M}, v \models \varphi \textit{ for some } v \in W$$

With E, it is possible to write a hybrid logic equivalent of $\langle R\rangle\varphi$. That is, $R \downarrow pE\varphi$ expresses that in the submodel restricted to Rw- worlds (marked by p), $E\varphi$ holds.

$$\mathcal{M}, w \models \langle R\rangle\varphi \Leftrightarrow \mathcal{M}, w \models R \downarrow pE\varphi$$

Last, the intersection modality introduced in section 4.1.3 is also necessary to describe concepts in game theory. The hybrid logic version of the intersection modality is:

$$\mathcal{M}, w \models \langle R \cap S\rangle\varphi \Leftrightarrow \mathcal{M}, w \models R \downarrow pS \downarrow qE((p \wedge q) \wedge \varphi)$$

The binding machinery from hybrid logic proves to be a valuable tool in efficiently describing game theoretic concepts such as best response, Nash equilibrium and strict domination.

Hybrid Logic Applied to Games

Recall that $\sim_i$, $\approx_i$ and $\succeq_i$ were originally referred to K_i, F_i and P_i in [40]. For the sake of readability, we will use these letters in the following definitions. Preference in hybrid logic encodes the notion that a player prefers a state which he can choose and is better than his other choices.

Fact 4.1.1.
$\mathcal{M}, w \models \langle\preceq_i\rangle\varphi \Leftrightarrow \mathcal{M}, w \models \langle F_i \cap P_i\rangle\varphi$
$\mathcal{M}, w \models \langle\prec_i\rangle\varphi \Leftrightarrow \mathcal{M}, w \models K_i \downarrow p\langle\preceq_i\rangle(\varphi \wedge \neg\langle\preceq_i\rangle p)$
$\mathcal{M}, w \models \langle\succeq_i\rangle\varphi \Leftrightarrow \mathcal{M}, w \models K_i \downarrow pE(\varphi \wedge \langle\preceq_i\rangle p)$
$\mathcal{M}, w \models \langle\succ_i\rangle\varphi \Leftrightarrow \mathcal{M}, w \models K_i \downarrow pE(\varphi \wedge \langle\prec_i\rangle p)$

The initial hybrid logic definition of $\langle\preceq_i\rangle$ postulates that preference is based on not only P, but also on F. One prefers a state if he can also choose it. Therefore, best response is a simple formula[5].

Definition 4.1.13. ***Best response*** *for i is true in a world w if*

$$\mathcal{M}, w \models [\prec_i]\bot$$

Van Benthem et al. (definition 6.1.9) also define best response approximately in this way: $\neg\langle\approx_i \cap \succ_i\rangle\top$. In fact, the above definition[6] can easily be derived from van Benthem et al.'s.

Of course, the definition of Nash equilibrium follows from best response ("w is a best response for every agent" [40]).

[5] With these relations as modalities, the global modality, etc., there are many ways (that amount to the same thing) to define solution concepts

[6] P should express strict preference in this case, i.e. $P(w, u, a, t)$ expresses at time t agent a regards world u better than w.

Definition 4.1.14. ***Nash Equilibrium*** *is true in a world w if*

$$\mathcal{M}, w \models \bigwedge_{i \in N} [\prec_i]\bot$$

With this new machinery it is also possible to define strictly dominated strategies. Seligman defines dominated strategies as $K_i \downarrow p\langle F_i\rangle[K_i]\langle\succ_i\rangle p$. However, this definition seems to leave open the possibility that the preference relation, instead of referring back to the worlds bound by p, refers to another world accessible by F that is also less preferred. Therefore, I propose that the following definition, where an additional binder ensures that the strict preference refers only back to the worlds bound by p.

Definition 4.1.15. *A strategy σ is* ***strictly dominated*** *if the following holds for every w in which i plays σ:*

$$\mathcal{M}, w \models K_i \downarrow pA(E(F \downarrow q\ \langle\succ_i\rangle(p \wedge q)))$$

With a definition for strict domination, we can consider characterizing IEDS in hybrid logic, for it has been shown by [13] and [2] to be a pivotal topic in game theory and modal logic. In order to do so, we must decide if it is possible to characterize rationality in hybrid logic.

"The Weak Rationality assertion WR_i was defined to fail exactly at those rows or columns in a two-player general game model that are strictly dominated for i" [2]. The hybrid logic definition of weak rationality is thus based on the following hybrid logic motivated formulation:

> A player j is weakly rational in a state ω if and only if for all $\approx_j$-accessible worlds υ from ω either ω is better than or equal to υ **or** there is a $\sim_j$-accessible world χ from ω that is better than or equal to some $\approx_j$-accessible state θ from χ.

Consider the following example[7]:

	A(j)	B	C
D(i)	3,2	2,1	1,1
E	2,1	0,0	0,1

Weak rationality holds for a state if the action is not dominated. Here we see that $\mathcal{M}, (D, A) \models WR_j$ since j plays A because it is equal to better than (for instance) C when i plays D, and it is better than B when i plays E. A hybrid logic definition for weak rationality is thus:

Definition 4.1.16. ***Weak Rationality*** *for player i in ω.*

$\mathcal{M}, \omega \models WR_i$ *iff* $\mathcal{M}, \omega \models F_i \downarrow pP_i \downarrow qE(p \wedge q) \vee K_i \downarrow rE(r \wedge F_i \downarrow sP_i \downarrow t(t \wedge u))$

[7] This example is a minor adjustment to an example in [2] and [4]

On the other hand, strong rationality can be based on the following formulation:

> A player i is strongly rational in state ω if and only if the current world ω is better than all F_i-accessible worlds υ from ω **or** some K_i-accessible world χ from ω is better than all F_i-accessible worlds θ from χ.

"Strong Rationality has a straightforward game-theoretic meaning: The current action of the player is a best response against at least one possible action of the opponent." [2] For the hybrid logic definition, we will need to make use of the universal modality:

Definition 4.1.17. *The **universal modality** A.*

$$\mathcal{M}, w \models A\varphi \text{ iff for all } v \in W, \mathcal{M}, v \models \varphi$$

Then the hybrid logic definition for strong rationality is:

Definition 4.1.18. ***Strong Rationality*** *for player i in ω:*

$$\mathcal{M}, \omega \models SR_i \text{ iff } \mathcal{M}, \omega \models F_i \downarrow pP_i \downarrow qA(p \wedge \neg q) \vee K_i \downarrow r(r \wedge F_i \downarrow sP_i \downarrow tA(s \wedge \neg t)$$

With the definitions of strong domination and rationality, it is possible to characterize IEDS in hybrid logic as well. This characterization is based on Bonanno's; if all players act (weakly) rational, then we result with a state where every player plays a strategy that is not strictly dominated. If $\mathcal{M}, w \models SD_i$ expresses that i is not playing strictly dominated strategy in w (and this holds for every w' such that wK_iw'), then the following characterizes IEDS:

$$\mathcal{M}, w \models \bigwedge_{i \in N} WR_i \Rightarrow \mathcal{M}, w \models \bigwedge_{i \in N} SD_i$$

4.1.5 Conclusion

This section has described various ways that modal logic can be used to describe concepts of strategic games. All the literature described here has suggested important factors to keep in mind when translating or creating new formulas in modal logic for strategic games. For instance, the literature helped us discern that the inherent presence of relations in strategic games are foundation to building an expressive logic. Last, we see that hybrid logic is a very powerful system that can describe game theoretic concepts efficiently and elegantly. The motivation behind introducing hybrid logic is that it will give us a way to define difficult concepts such as evolutionary stable strategy in evolutionary game theory. This will be thoroughly explored later on. In the following section, I will describe logic for extensive form games. Subsequently, I will compare the salient themes in both strategic and extensive form games.

4.2 Logic for Extensive Form Games

In the previous section on logic and strategic games, we saw that relations together with hybrid logic were appealing for describing strategic games. This section concerns extensive form games, which, as we saw in chapter 2, describes sequential games and game trees. An extensive form game is closely related to a strategic game, because the strategies and preferences allow us to translate an extensive game into a strategic one by means of normalisation. The only problem with doing that is that even though we still have all the outcomes represented accurately, we sacrifice the intricate nature of a sequential game, which in many ways encode the rational deliberation of players.

The literature I surveyed in section 4.1, however also thoroughly discuss extensive form games. In this section, I will therefore also explain what those articles say about extensive form games. As the iterated eliminated of dominated strategies was an important theme in strategic form games, backward induction reflects the same kind of relevance for us in extensive form games.

First, in section 4.2.1, I will briefly describe how Bonanno applies the relation R from his second view of game theory (prescriptive) to describe backward induction. It's probably also prudent to explain why only that interpretation of R works for BI.

After describing Bonanno's account of BI, I will, in section 4.2.2, turn to van Benthem's theory of play, which also describes some interesting approaches to backward induction, because logicians are interested in modelling strategic interaction involving rationality, actions, and preferences with logic. This section will focus on two specific of ways of thinking about backward induction; first in terms of the logic of extensive form games as process models, and the other in terms of the forcing relation connecting strategies to outcomes from a specific node in a game.

In the following section, section 4.2.3, I propose a new perspective on the relation freedom in extensive form games, because freedom does not make sense for extensive games. The relations are unrealistic given the sequential nature of extensive form games. I propose that we must instead define freedom as a relation over sets of outcomes and not outcomes individually. In this way, the freedom relation expresses exactly what it does in strategic games but with a slightly different formulation. In this section I will also suggest that if we the factor of each player's preferences, then it again becomes possible to make freedom (actually freedom$^+$) relations between individual outcomes. This concept allows us to define a backward induction solution using the freedom relation.

Section 4.2.4 will conclude this discussion of logic and extensive form games with 1.) how we can describe process models in standard game theoretic terms, and 2.) a comparison of forcing relations/strategic powers, which were outlined in section 4.2.2, and my own contribution regarding freedom in section 4.2.4. This will address the close relationship between freedom and forcing/strategic powers.

4.2.1 Bonanno's Account of Backward Induction

Recall that Bonanno distinguished two interpretations of game theory: as a description of rational behaviour and a recommendation to players on how to act. The latter view changed the interpretation of the relation between worlds, $\alpha R_i \beta$, from "for player i, β is epistemically accessible from α" to "from state α, player i can unilaterally bring about state β." With the latter interpretation of R, Bonanno characterizes a game model based on the notion that solution concepts are motivated by *recommendation.* Since R describes a players ability to bring about a node unilaterally, thus giving him the ability to follow the theory's recommendation, which is represented by R_* (if $\alpha R_* \beta$, then it is recommended at state α that state β be reached). The main result from Bonanno's model based on recommendation is a characterization of the backward induction algorithm.

> ... the backward-induction algorithm determines for every decision node a unique immediate successor, thus giving rise to a relation on the set of nodes Ω. Call it the backward-induction relation. We say that the relation R_* is the backward induction recommendation if it is the transitive closure of the backward-induction relation. [13]

Freedom is obviously the relation at issue in extensive form games for Bonanno since extensive form games fall under the second interpretation of game theory where R is freedom.

4.2.2 The Theory of Play and Extensive Form Games

A large and comprehensive amount of information about modal logic and the forms of classical game theory has been described by "The Theory of Play" article by Johan van Benthem, Eric Pacuit and Olivier Roy[10], and van Benthem's upcoming book *Logic in Games* [4]. The *theory of play* aims to describe the full process of a game; the expectations before the start of a game, the reasoning and deliberation during a game, and the rationalization of the game afterwards. *Logic in Games*, which elaborates on the theory of play, also takes a large step back to describe the general picture of the interaction of logic and game theory. From this broader perspective, van Benthem describes extensive form games as process models (and describes it with a process logic), which is an interesting revelation on how "logic clarifies relevant process structure" [4]. This section will focus on two modal logic interpretations of extensive form games and backward induction.

First I explain the process models and logics, which describe game trees as "a process with states and transitions..." [4]. In the theory of play, the backward induction algorithm is described in terms of a modal logic based on process models, which "[combines] the logics of action and preferences" [10]. Backward induction is considered to be a "pilot example" of the interaction of rationality and actions.

This view of extensive form games entails an interesting feature of processes, *forcing*, which will be the second feature of game play this section considers: the "potential interaction" that can be observed at a point in the game, where it can be postulated which set composes the possible outcomes resulting from that point given a player's strategy. I consider this an interesting example of modal logic of game theory; this is because after I explain how Seligman's relation's (in particular, freedom) operate in extensive form games, I will demonstrate that has clear and informative connections with this notion of forcing.

Extensive Form Games as Process Models

Another way one can describe extensive form games as a "model for a modal language" [4] is by considering them as process models. With the language of the process graphs, the *BI* algorithm is defined in a logical manner. The general *BI* algorithm is:

> At each stage, mark the dominated moves in the $\forall\forall$ sense of set preference as 'passive', leaving all others active. In this comparison, reachable endpoints by an active move are all those that can be reached via moves that are still active at this stage. [4]

Where the $\forall\forall$ stipulation is a view of propositional preference that claims that a set Y is preferred to a set X if all members of Y are better than (or equal to) all members of X; that is, Y is preferred to X if $\forall x \in X\ \forall y \in Y\ x \leq y$. The following language, which is to describe extensive form games as processes, operates as a model for modal logics.

Definition 4.2.1. ***Extensive Form Game.***
An extensive form game is a tree $\mathcal{M} = \langle \mathsf{NODES}, \mathsf{MOVES}, \boldsymbol{turn}, \boldsymbol{end}, V \rangle$ *with binary transition relations from the set* MOVES *pointing from parent to daughter nodes. Non-final nodes have unary proposition letters* $\boldsymbol{turn}_i$ *indicated the player whose turn it is, while* ***end*** *marks end nodes. The valuation V can also interpret other local predicates at nodes, such as utility values for players or more ad-hoc properties of game states. [4]*

We use *bi* to denote the subrelation of the total *move* relation produced by the algorithm. A general formulation of the *BI strategy*, appropriate for logics of action is:

> The *BI* strategy is the unique relation σ satisfying the following modal axiom for all propositions p – viewed as sets of nodes – for all players i [4]:
>
> $$(turn_i \wedge \langle best \rangle [best^*](\mathbf{end} \to p)) \to [move_i]\langle best^* \rangle (\mathbf{end} \wedge \langle pref_i \rangle p)$$

Where $move_i = \bigcup_{a \text{ is an } i \text{ move}} a$ and $turn_i$ denotes that it is i's turn to move, and *end* is a propositional variable true at end nodes [10].

With the logic that describes extensive form games as process models, it is possible to describe the fact that at the current node, there is a strategy for player i, which in response to j's initial move results in a φ–state after two steps of play [4] with the following modal formula: $[move - i]\langle move - j\rangle\varphi$.

This modality expresses that j has a strategy in responding to i's move that results in an outcome where φ holds. This is the notion of a player's strategic powers which is described in the next section.

Players' Strategic Powers

van Benthem in [4] introduces a way of observing the process of the game, where each node is related to some outcomes given one player's strategy. It describes the extent to which the players have the power to achieve particular outcomes: "games are all about powers over outcomes that players can exercise via their strategies" [4]. I will later argue that there is a clear relationship between the concept of strategic powers and my application of freedom to extensive form games.

The basic idea of a player i's strategic powers is that at a state s in a game, i can force a set of outcomes X when for all of his opponents' strategies, if i follows a particular strategy, it results in a member of X. This creates a relationship between the state s and a set of outcomes. With a modality, we can be more specific and define the relationship between the state s and the outcomes where some formula φ holds as a forcing modality:

Definition 4.2.2. ***Forcing modalities*** $\{i\}\varphi$.
$\mathcal{M}, s \models \{i\}\varphi$ *iff player i has a strategy for the subgame starting at s which guarantees that only end nodes will be reached where φ holds, whatever the other player does.*

Moreover, we can use this notion of a forcing modality to establish a set based on the fact that a player playing a strategy starting a state s: the resulting outcomes according to that strategy compose a set X.

Definition 4.2.3. *The* ***forcing relations*** $\pi_i^\Gamma(s, X)$ *in a game tree hold if player i has a strategy for playing game Γ from state s onward whose resulting states are always in the set X. When s is known in context (often it is the root), the sets X are called the powers of player i.*

In the game depicted in figure 4.3, player 2 has a strategy by which he can force a state where q is true. Formally, $\mathcal{M}, s \models \{2\}q$.

A broader view of strategic powers involves naming the full sets of possible outcomes that result from some state onwards when i follows the strategy σ. For every strategy σ listed for a player, there is a corresponding set resulting from some state onward. The larger a set is, the weaker the power is. The following properties formally describe properties of players' powers:

Closed Under Supersets If $\pi_i^\Gamma(s, Y)$ and $Y \subset Z$, then $\pi_i^\Gamma(s, Z)$

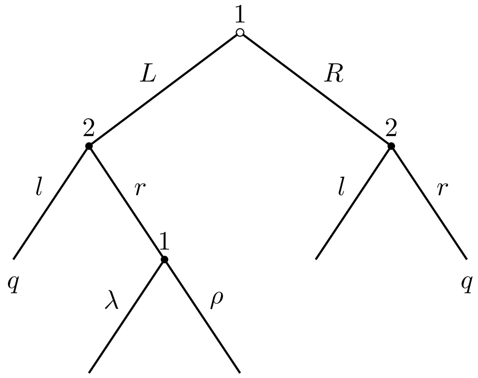


Figure 4.3

Consistency If $\pi_1^\Gamma(s, Y)$ and $\pi_2^\Gamma(s, Z)$, then Y, Z overlap[8]

Completeness If not $\pi_1^\Gamma(s, Y)$, then $\pi_2^\Gamma(s, S \backslash Y)$ and vice versa, where S is the total set of outcome states.

Forcing and players' strategic powers will be come up again in section 4.2.4 as a feature of extensive form games that are closely related to the proposed definition of freedom for extensive form games.

4.2.3 Freedom in Extensive Form Games

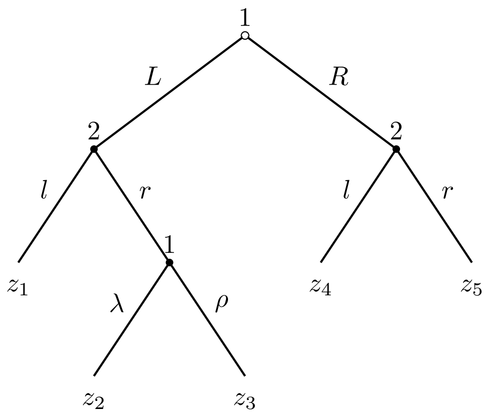
Now that we have seen a few approaches to logic and extensive form games in Bonanno's and van Benthem's work, we turn to Seligman's relational model with freedom, knowledge and preference. This section focuses on the relation freedom, because the notion of choice is central to this thesis; moreover, knowledge is an already thoroughly studied feature of games. Nevertheless it still holds that $\sigma \approx_i \sigma'$ means $\sigma \sim_{-i} \sigma'$. This equivalence is also still crucial to understanding how a game is structured.

Defining Freedom for Extensive Form Games is Problematic

The relationships that were demonstrated in strategic games do not translate to extensive form games straightforwardly. The relation $\approx$ does not behave the same in extensive games, and therefore I will seek a new definition for freedom that fits game trees without losing its original meaning. Following this, I will demonstrate that there is a connection between the theory of play and the new formulation of freedom that fits for extensive form games.

[8] "... players cannot force the game into disjoint sets of outcomes, or a contradiction would result..." [10]

We know that every extensive form game has a corresponding strategic form game through normalisation. As seen in section 4.1, relationships such as freedom, uncertainty and preference exist between the outcomes represented in a strategic form game. Given the connection between extensive form games and strategic form games, we should be able to identify the same relationships in an extensive form game from which a corresponding strategic form game originated by normalisation. Preference is already included in the standard extensive form game $\Gamma = \langle N, H, P, (\succeq_i)_{i \in N} \rangle$.



(a) Extensive form game

	ll	lr	rl	rr
$L\lambda$	z_1	z_1	z_2	z_2
$L\rho$	z_1	z_1	z_3	z_3
$R\lambda$	z_4	z_5	z_4	z_5
$R\rho$	z_4	z_5	z_4	z_5

(b) Normalized strategic form game

Figure 4.4

Consider the game tree and game matrix in figure 4.4. According to the reasoning from Section 4.1, the follow relations hold for the strategic form game in that figure:

$$\approx_1 = \{(z_1, z_4),\ (z_1, z_5),\ (z_2, z_3),\ (z_2, z_4),\ (z_3, z_4),\ (z_2, z_5),\ (z_3, z_5)\}$$

$$\approx_2 = \{(z_1, z_2),\ (z_1, z_3), (z_4, z_5)\}$$

Figure 4.5 depicts the extensive form game from figure 4.4a with the dotted lines to represent the freedom relations that hold for player 1 in the strategic form game in figure 4.4b.

Because extensive form games represent sequential turn-taking games, it is difficult to justify some of the above relations. For instance, according to the set

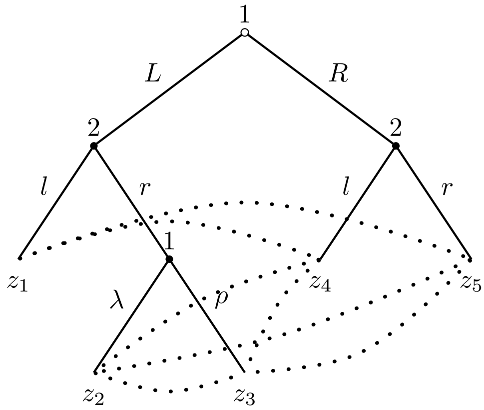


Figure 4.5: The dotted lines represent the freedom relations for player 1 as suggested by the standard strategic definition of freedom.

$\approx_1$, player 1 has the freedom to choose between outcomes z_1 and z_4. But player 1 does not have the freedom to choose between z_1 and z_4, because z_1 and z_4 are terminal nodes resulting directly from actions available at histories where it is player 2's turn to play. At best, player 1 can "set up" player 2 to play either at the history where he could choose z_1 or the history where he could choose z_4[9].

We can conclude from this objection that player 1, in fact, has freedom over sets of outcomes in $\approx_2$ and not freedom over the outcomes as listed above according to the reasoning from 4.1. A player, first and foremost, at different nodes in a game tree has a *choice* and thus freedom over *something*.

At player 1's first decision node in our example, he can choose which choices player 2 will have (in other words, player 1 chooses player 2's first decision node). Player 1 therefore chooses between the following two possibilities:

- outcome z_1 or another choice node for himself,
- outcomes z_4 or outcome z_5.

It would be nice to express this informal idea about freedom over *sets* of outcomes in the standard extensive form game terminology.

A Modified Definition of Freedom

Recall the terminology regarding the structure for extensive form games $\Gamma = \langle N, H, P, (\succeq_i)_{i\in N}\rangle$ as well as the specific terminology regarding subgames where a subgame of Γ is $\Gamma(h) = \langle N, H|_h, P|_h, (\succeq_i |_h)\rangle$. The objective is to define the freedom relation between sets of outcomes by limiting the set of all terminal histories of Γ to a set of terminal histories $Z|_h \subseteq Z \subset H$ that follow from one

[9]Only because those histories result from a history at which it is player 1's turn to play. As a foreshadowing to the reader, this line of reasoning will lead to an alternative suggestion for how to think about freedom in extensive form games.

particular history h. We can restrict the set of terminal histories to the ones that result from one history h by referring to the subgame[10] of Γ induced by h, or $\Gamma(h)$.

Definition 4.2.4. *The **set of terminal histories** resulting from a history h' in the subgame $\Gamma(h)$ is*

$$Z|_h(h') = \{z : (h, h') = z \text{ for every } h' \in H|_h\}$$

Freedom for a player is a relation over such sets. We thus define[11] freedom as follows:

Definition 4.2.5. ***Freedom for i in Extensive Form Games.***
In a subgame $\Gamma(h)$, $h \notin Z$, of Γ where $P(h) = i$, $Z|_h(h') \approx_i Z|_h(h'')$ if and only if:

1. *$a', a'', ... \in A(h)$ such that $(h, a') = h'$, $(h, a'') = h''$ ..., and*
2. *If a history $h' \notin H|_h \cap Z$, then for all $a \in A(h')$, $(h', a) \in Z|_h(h')$.*

This definition encodes that a player i has the freedom to choose between sets of outcomes[12] if the subgames resulting from the actions he may choose from result in those outcomes, or another history resulting in eventual outcomes.

At subgame $\Gamma(L)$ in 4.6, the histories $h' \in H|_L$ are $\{L, Ll, Lr, Lr\lambda, Lr\rho\}$. In this subgame the sets of terminal histories $\{z_1\}$ and $\{z_2, z_3\}$ satisfy the above conditions for freedom for player 2; that is, $P(L) = 2$, the actions available to 2 are l and r, where $(L, l) = z_1$ and $(L, r) = h_2$. One history is not in the set of terminal nodes, $Lr \notin Z$, but that history satisfies the second condition that all actions available at Lr, which are λ and ρ, result in the terminal nodes $z_2, z_3 \in Z|_L(Lr)$. Thus, $Z|_L(z_1) \approx_i Z|_L(Lr)$.

We can, however, get more specific about what a player thinks he is choosing between. With the preference relation we can achieve this.

Freedom$^+$
Preference simplifies the many choices with which players are confronted, and it is essential to making a game a real game: "available actions and information

[10] As a reminder to the reader, (h, h') is just a sequence of actions described by the actions in h together with the actions in h'

[11] In words, what this definition should express is: In a subgame $\Gamma(h')$ of Γ where it is i's turn, i has the freedom to choose between two sets of outcomes $Z|_h(h')$ or $Z|_h(h'')$ if and only if

1. The actions $a', a'', ...$ available to i at h result in histories $h', h'', ... \in H|_h$, respectively, and
2. Either $h', h'', ... \in H|_h \cap Z$ or, if not, for histories $h', h'', ... \in H|_h$ are not terminal nodes, but the actions resulting from that history is in the set of terminal nodes of the subgame starting at the relevant history $h', h'', ...$, or in other words the relevant $h', h'' \in Z|_h(h')$.

[12] Possibly singletons

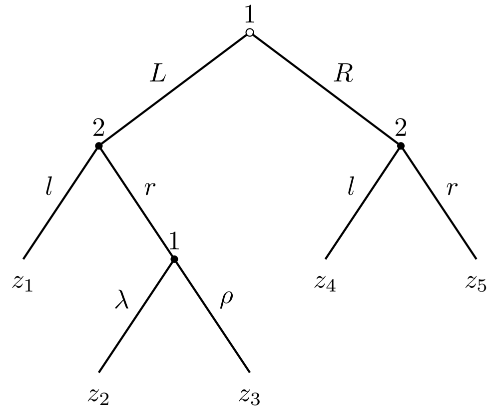


Figure 4.6

give the 'kinematics' of what can happen in a game – but it is only their interplay with evaluation that provides a more explanatory 'dynamics' of well-considered intelligent behaviour" [4].

In extensive form games, if $Z|_h(h') \approx_i Z|_h(h'')$, then it follows that i has no "control" over which of the members of these sets of terminal nodes will actually result. However, if player i knows his opponent's preferences and that he is rational, player i has much more awareness over the situation. Player 1 can restrict his opponent's next choice by means of what choice his opponent prefers. Consider a game Γ where:

- $P(h) = 1$
- $Z|_h(h') = \{z_1, z_2\}$, and $z_2 \succ_2 z_1$.
- $Z|_h(h'') = \{z_4, z_5\}$ and $z_4 \succ_2 z_5$.

Given that player 1 knows the preferences and knows that player 2 is rational, he is in fact choosing between outcomes z_2 and z_4 *instead* of between the sets $\{z_1, z_2\}$ and $\{z_4, z_5\}$, because player 2 will not play any of the other actions leading to z_1 or z_3. In other words, player 1 has freedom over player 2's *preferred* outcomes. We will continue to refer to this as freedom$^+$ or $\approx_i^+$. Thus, $z_2 \approx_i^+ z_4$.

Definition 4.2.6. ***Freedom***$^+$
$z' \approx_i^+ z''$ if and only if for $z' \in Z|_h(h')$ and $z'' \in Z|_h(h'')$ where $Z|_h(h') \approx_i Z|_h(h'')$ the following holds: for all $z \neq z' \in Z|_h(h')$, $z' \succ_{-i} z$ and for all $z \neq z'' \in Z|_h(h'')$, $z'' \succ_{-i} z$.

Given this definition, where a player i is able to choose among two outcomes z' and z'' that have survived the preferences of each player in the progression up the game tree, all that is left is for i to act upon his preferences and take which ever action that will result in the best outcome for him. If $z' \approx_i^+ z''$ and $z' \succ_i z''$ then z' is the *Backward Induction solution* for this game.

Definition 4.2.7. *The outcome z' is the* ***Backward Induction solution*** *for a game Γ if and only if for all $z'' \in Z$ such that $z' \approx_i^+ z''$, $z' \succ_i z''$.*

Example
Consider the following example in figure 4.7 as an illustration of freedom and freedom$^+$.

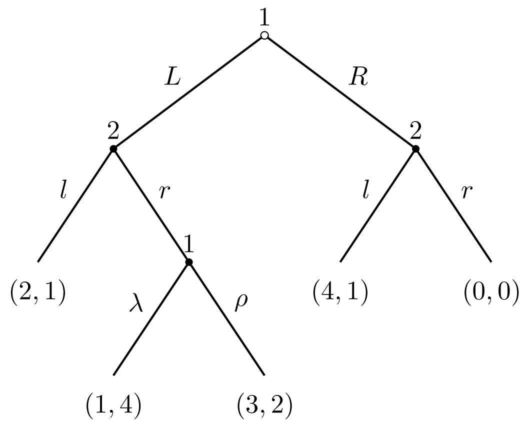


Figure 4.7

- Player 1 has freedom over the following sets:
  $Z|_\emptyset(\emptyset L) \approx_1 Z|_\emptyset(\emptyset R)$ and $Z|_\emptyset(\emptyset Lr\lambda) \approx_1 Z|_\emptyset(\emptyset Lr\rho)$, which means that he can choose between sets of outcomes $\{z_1, \{z_2, z_3\}\}$ and $\{z_4, z_5\}$ and between $\{z_2\}$ and $\{z_3\}$.
- Player 2 has freedom over the following sets:
  $Z_{\emptyset L}(\emptyset Ll) \approx_2 Z_{\emptyset L}(\emptyset Lr)$, which means that he can choose between sets of outcomes $\{z_1\}$ and $\{z_2, z_3\}$ and between $\{z_4\}$ and $\{z_5\}$.

We saw in 4.2.2 that for sets of players' powers, "the bigger the set, the weaker the power." Here we can also claim that a bigger set implies weaker choice over outcomes. Given that some powers are weaker than others, [4] claims that we can drop the sets for which there is a stronger set. In figure 4.7, player 2 has, for instance, the following powers: $\{z_1\}$ and $\{z_1, z_5\}$. Because $\{z_1\}$ is stronger, $\{z_1, z_5\}$ can be dropped.

Along similar reasoning, we can conclude that because player 1 has freedom to choose between $\{z_2\}$ and $\{z_3\}$ and between $\{z_1, \{z_2, z_3\}\}$ and $\{z_4, z_5\}$, and $\{z_1, \{z_2, z_3\}\}$ is weaker than $\{z_2\}$ and $\{z_3\}$, that therefore 1 can actually choose between $\{z_1, z_2\}$ and $\{z_4, z_5\}$ and between $\{z_1, z_3\}$ and $\{z_4, z_5\}$.

4.2.4 Freedom, Forcing and the Process Model

This section describes the connections between freedom as described in section 4.2.3 and the notions of freedom and forcing.

Forcing and Freedom

I demonstrated in the previous section 4.2.3 that freedom for a player i means that i has can choose between the sets of outcomes that are composed of individual outcomes that result from each of his possible actions at a history h. The forcing relations, on the other hand, results in a set out of outcomes that result from a state s via some strategy σ_i for i. For consistency, we will at this point to think of the forcing relations of i as sets resulting from a history h instead of a state s.

Some initial similarities and differences of powers and freedom can be determined from the get-go:
First, powers resemble the knowledge relation $\sim$: There is a one-one relationship between a player's possible strategies in an extensive game and his powers. Moreover, the powers represent the sets resulting from every strategy of a player *without the opponent's strategy accounted for*. Player 1 does not know what player 2 will do, so player 1 only knows that the outcome will be a member of the 'power' sets corresponding to each of his strategies. This indicates a relationship similar to knowledge, because only a player's opponent's strategy is a variable.

Second, forcing relations are bound to a history h and some strategy σ, whereas freedom is just bound to a history h. It follows that:

- Freedom for i runs between sets which are composed of *all* possible outcomes resulting from h via the actions in $A(h)$.
- Powers are the sets of *some* outcomes resulting from a history h. The "some" is based on the fact that these sets result from the strategy starting at history h and follow *one* strategy for i, regardless of the opponent's moves.

Consider the following figure 4.8. This example will demonstrate a consequence of the above two fact.

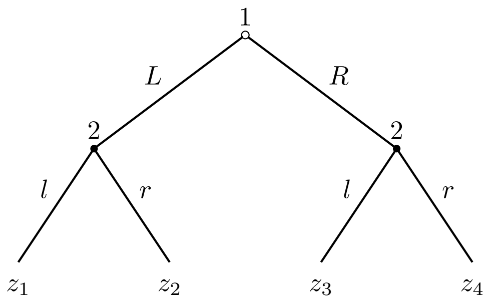


Figure 4.8

Consider the extensive form game in figure 4.8. Player 2 has the freedom relations $z_1 \approx_2 z_2$ and $z_3 \approx_2 z_4$. For now, let us also refer to these as "freedom

sets": $\{z_1, z_2\}$ and $\{z_3, z_4\}$.
The forcing relations, on the other hand result in strategic "power sets". In figure 4.8, player 2 has the forcing sets, $\{z_1, z_3\}, \{z_1, z_4\}, \{z_2, z_3\}$, and $\{z_2, z_4\}$. It is apparent at this point that forcing and freedom do not from a state h result in the same sets.

However, suppose that the game depicted in 4.8 is the subgame of a larger game, where the history labelled $\emptyset$ is results from some action taken at the previous step in the game, such that $\emptyset \in A(\emptyset - 1)$ where (as we know from the figure) $P(\emptyset) = 1$ and $P(\emptyset - 1) = 2$. It follows that the subgame starting at $\emptyset$ leads to terminal nodes which are one member of player 2's greater freedom relations. That is, $Z|_{\emptyset-1}(\emptyset) \approx_2 Z_{\emptyset-1}(h)$ for some h such that $(\emptyset, h) \in H|_\emptyset$

Process Model

Note that these definitions reflect [10]'s descriptions of BI and rationality. The notions in these definitions are describable in standard game theoretic terminology, including the new terms defined above such as $Z|_h(h')$, freedom, and freedom$^+$. For instance, the BI scenario we are concerned with "at each stage, mark[s] dominated moves in the $\forall\forall$ sense of preference as *passive*, leaving all others active" [10] can be defined as:

> A move a is *passive* at a stage h where $P(h) = i$ if $\forall z' \in Z|_h(h')$ and $\forall z'' \in Z|_h(h'')$ for all $a' \neq a \in A(h)$ such that $(h, a) = h'$, $z'' \succ_i z'$

And "here 'reachable end-points' by a move are all those that can be reached via a sequence of moves that are still active at this stage" [10]. The reachable end-points by a move ($a \in A(h)$) is exactly what a set $Z_h(h')$ stands for, where the members of that set are not passive[13].

Figure 4.9 fits the $\forall\forall$ picture:

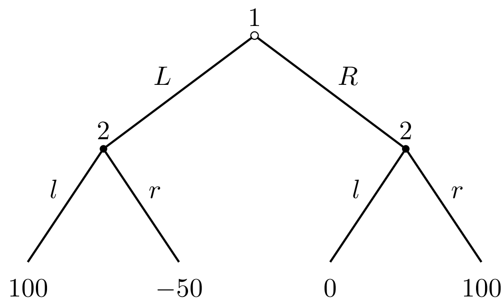


Figure 4.9: This is an extensive form game with strategies for player 1, strategies for player 2 and values for outcomes.

[13]One would be able to define active in a similar way to how passive was defined above.

At the initial node, the move R is marked passive, because all outcomes (end-points) resulting from move R are preferred by player 1 to those resulting from move L. Therefore outcomes with utilities $(3, 1)$ and $(3, 2)$ remain active. By the end of the game, only (3,2) remains active, for player 2 prefers that outcome of move r to the outcome of move l.

This example works because it is ideal, but many games will have outcomes that do not allow a player to prefer *all* outcomes of one move to *all* outcomes of another move. This encourages us to look at an alternative notion of preference to the ∀∀ notion.

4.3 Conclusion

The goal of this chapter was to survey existing contributions towards the field of classical game theory and logic as well as to introduce some novel proposals that shed some light on the theories explained in the survey. This topic is easily split in two, where one part describes logic for strategic form games, and the other part describes logic for extensive form games.

Section 4.1 explored the logics for strategic form games. Logics for game matrices are arguably less popular than the logics for extensive form games, for it is not immediately obvious where the modalities "occur." Nevertheless, I wish to emphasize the value of the strategic form game, even in the light of modal logic; after all, as Seligman pointed out in [40], there exist some very interesting and subtle modalities that elegantly describe the player's relationship to the game, and the matrix game's structure itself. To take it a step further, we saw how easily hybrid logic applied to strategic form games. We saw that the important solution concept in strategic form games, IEDS, could easily be described by hybrid logic as well.

Section 4.2, on the other hand, described the more obvious connection between logic and extensive form games. An extensive form game naturally fits the classic idea modal logic. It is, after all, a model with states and connections to states and outcomes. Modal logic is also an appealing formal framework for extensive form games, because both concern the rational interaction of agents. Epistemic logic, a modal logic, is therefore superbly equipped to describe extensive form games.

This section described modal logic approaches to extensive form games from the sources we saw in the survey in section 4.1. First, Bonanno's account of backward induction was briefly outlined. Following this, some aspects of the theory of play were examined; process models and strategic powers. As in the other sections, the section is wrapped up with a novel analysis of Seligman's freedom relation under extensive form games and how the view connects with the existing material on logic and extensive form games.

In the next chapter, the discussion will turn to evolutionary game theory. I will contend that it is not as straightforward as to simply apply the logics seen in classical game theory to evolutionary game theory. Instead, the chapter will begin with a discussion of the role of rationality in evolutionary game theory. The changing role of rationality as well as the reinterpretation of classical terms, as described in chapter 3, will prompt the need for a different perspective for the logic of evolutionary game theory.

Chapter 5

Logic in Evolutionary Game Theory

The game logic frontier is a relatively new field, and evolutionary game theory itself is also relatively new. Therefore, there is also relatively little written about the relationship between logic and evolutionary game theory. The goal of this chapter is to take some initial steps towards an evolutionary game logic. Recall the thematic statement by John Maynard Smith in his paper "Evolutionary Game Theory" [41]:

> *There are two main differences between classical and evolutionary game theory.*
>
> 1. *The replacement of "utility" by "fitness"...*
> 2. *The replacement of rationality by natural selection.*

As described in chapter 3, the machinery in evolutionary game theory is the same as in classical game theory, but the difference lies in the interpretation of the terminology as indicated by Maynard Smith above.

The goal of this chapter is two-fold. I have argued in this thesis that the replacement, or purging, of rationality in evolutionary game theory is problematic, because game theory is generally reliant on rationality. Rationality has been seen in classical game theory and the corresponding game logics as a "bridge law" connecting information, preferences and actions. Rejecting rationality would seem to leave the connections between those concepts wanting. Therefore, I will investigate if an alternative view of rationality is possible under the evolutionary view. I propose three alternatives, which turn out to be as insufficient as the original view of rationality. This part of the chapter also describes how linguists have applied both classical and evolutionary game theory to describe pragmatics in natural language. This study demonstrates that a form of game theory without rationality can useful in applications.

The second aim of this chapter addresses the challenges of figuring out an evolutionary game logic. I will first describe why it is (close to) impossible to define evolutionary stable strategies in terms of the (relational) modal logic of the theory of play. Subsequently, I will propose two ways to define evolutionary stable strategy by means of hybrid logic, where one way is superior to the other in how intuitively it describes evolutionary game theory.

5.1 The Problem of Choice and Rationality

Evolutionary game theory, by nature, purges rationality. Classical game theory is founded on rationality, and many of the results are fully dependent on it:

> The models we study assume that the decision maker is "rational" in the sense that he is aware of his alternatives , forms expectations about any unknowns, has clear preferences, and chooses his action deliberately after some process of deliberation. [35]

> Game theory has originally been conceived as a theory of strategic interaction among fully rational agents ... Rationality here means, among other things, full awareness of ones own beliefs and preferences and logical omniscience. Even stronger, for classical game theory to be applicable, every agent has to ascribe full rationality to each other agent. [29]

> ... players never choose an action whose outcomes they believe to be worse than those of some other available action ... [4]

An interesting exercise, on the other hand, is to ask 'what classical game theory would look like without rationality?' According to van Benthem,"the role of rationality [is] a 'bridge law' between information, action and preference..." [4]. Rationality[1] is important to justify the link between the *utility* of an outcome and executing an *action* that results[2] in that outcome.

Because evolutionary game theory is composed of the same machinery as (but different interpretation of) classical game theory, we are faced with the challenge of evaluating the bare bones of the model without the meat (rationality) to hold it all together. Many descriptions of evolutionary game theory begin with the "disclaimer" that the the concept of rationality which is so central to game theory is lost under the evolutionary interpretation.

> [...] really stupid critters can evolve towards the solution of games previously thought to require rationality. [23]

[1]The above quote by [4] continues to point out that "[rationality as a bridge law] is packed with assumptions, and logic wants to clarify this, not endorse it."

[2]The definition of *rationality* is vague and there exist many diverse views on exactly what it means, but perhaps it should be made explicit here that rationality also implies *intelligence*. If rationality does not imply intelligence, then even if a player knows and intends to play the best outcome, it does not imply that he has the mental capacity to reason about his opponent. Nash equilibrium is dependent on the fact that each player can reason about the other and choose his best response accordindly.

> The evolutionary interpretation of game theory [...] completely gives up any rationality assumptions. [29]

But this section will not suggest that evolutionary game theory is void of a 'bridge law' connecting action and preference, but it will claim that because we have reinterpreted standard game theoretic concepts, we must also reinterpret the concept rationality in evolutionary game theory. We will consider some familiar discussions including those of Bonnano[13], and van Benthem[2], but also views from van Benthem and Jan van Eijck [7] and Herbert Gintis [23].

5.1.1 Alternative Views of Rationality

Instead of abandoning rationality completely, we can explore a few alternative ways to better make sense of rationality under the evolutionary interpretation. As mentioned above, van Benthem views rationality as a bridge law between actions, information and preferences. I will argue here that this bridge law can take different forms. Here I describe three: "as if"–rationality, where their behaviour mimics normal rationality, rationality as following a recommendation, and actions as revealing preference instead of following rationality. However, each of these alternative views of rationality have drawbacks, which I will also describe. Concluding this section, I will make a case for a possible solution to the problem, that is not an alternative view of rationality, but rather a way in which the abandonment of it is acceptable for game theory.

"As if they are rational"

A simple solution would be to view that bridge law, not as rationality, but as "as if"–rationality [6]. It is easy to anthropomorphize processes involving in natural selection; fish "choose" to swim in a school instead of alone, because they "know" that it makes it easier to avoid being eaten. This is a backwards way of reasoning about the advancement of a strategy in a population, for "... a gene is just a molecule, it can't *choose* to maximize its fitness..." [11]. After all, the actual scenario is that fish swim in schools, because the ones who did not (who were programmed with the instinct to swim alone) got eaten by bigger fish, leaving only the more successful fish programmed to swim in schools to replicate. So with regards to the fact that a gene cannot choose to maximize its fitness, "evolution makes it seem as though it had" [11].

Ken Binmore in [11] suggests that "as if"–rationality is very useful for biologists who have extremely complex chemical and physical reasons for calculating why one genetic instinct is better than another. The appeal of this non-rational rationality is it gives us a straightforward and intuitive explanation of the causes behind natural selection. After all, fish "neither know nor care that this behaviour is rational. They just do what they do. But the net effect of an immensely complicated evolutionary process is that [fish] behave *as though* they had chosen to maximize their fitness" [11].

With this "as if"–rationality, there is no problem preserving the bridge law between information, preferences and actions. The drawback of this view is that it sweeps the problem under the rug. The evolutionary process still has nothing to do with information or preferences. Only actions are relevant, and they are programmed behaviours; never choices.

Game Theory as Description or Prescription

Bonanno, in [13], claims that there are at least four distinguishable views of game theory. These four views are (1) "a *description* of how *rational* individuals behave," (2) "a *prescription* or *advice* to players on how to act," (3) an observation of how individuals actually act, and (4) in evolutionary terms, where "outcomes are explained in terms of dynamic processes of natural selection." Noticeably, Bononno lists evolutionary game theory as a separate view of game theory, rather than a reinterpretation or extension of it.

Nevertheless, this thesis has addressed what is actually a descriptive look at evolutionary game theory, so it must also be possible to view evolutionary game theory in the light of the prescriptive view. This would entail that a strategy would be recommended to a player based on how well it will do in a population. A player would get a new recommendation after every encounter with an opponent, based on the frequency of the strategies. But solution concepts can never be prescriptive, for they are judged by behaviour over time, and a recommendation can only be made on individual encounters of players.

Preference Revelation

van Benthem and van Eijck suggest in [7] that the actions taken by players in a game are not inherently rational or irrational, because there is always a way of assigning preferences to the outcomes of a game to *make* it so that the actions taken were, in fact, rational. And therefore, it can be interpreted as preference revelation: "Instead of condemning a particular move as irrational, one might wish to take the move as a revelation of an agent's preference" [7].

I argued in the "as if"–rationality section that "It is easy to anthropomorphize processes involving in natural selection; fish 'choose' to swim in a school instead of alone, because they 'know' that it makes it easier to avoid being eaten. This is a backwards way of reasoning about the advancement of a strategy in a population." The suggestion by [7] of after-the-fact rationality (similarly backwards) is an appealing way of thinking about the concept for biologists. They are, by definition, always after-the-fact rationalizing the behaviour of animals, for they can only observe animals' choices and what the outcome of those choices were. Only then can we conclude what the utility of that strategy profile was. Moreover, the frequency of a strategy in a population changes over time, so the observer is getting new information about actions, utilities and therefore new preferences of a player after each encounter.

This seems to be the only viable view of rationality for evolutionary game theory, but this is not exactly the rationality that is required for classical game theory. So we can attribute this “version” of rationality to evolutionary games, but the problem of the abandonment of the rationality in classical game theory is still there.

5.1.2 Rationality in Game Theoretic Pragmatics

Through the application of language to classical and evolutionary game theory, we can make a case study of the role of rationality in each theory.
Language is a phenomenon that lends itself to an evolutionary game theoretic interpretation. Many components in language ranging from the physical ability of hearing to language acquisition can be seen in an evolutionary light [28]. Although this thesis has posed the “loss” of rationality as a big obstacle, the fact that rationality plays a weak role in evolutionary game theory is under some circumstances favourable. For instance, the pragmatic coordination of meaning as described by David Lewis' [31] signalling games have primarily been explained by means of classical game theory with the standard role of rationality. With evolutionary game theory, on the other hand, “in order to successfully communicate information, we don't need as much rationality, higher order intentionality or common knowledge (explicitly or implicitly) required by Grice ... and others” [38].

Rationally Justified Conventions in Signalling Games

Communication can be modelled as a signalling game, which is is a two player game describing the strategic coordination of information exchange. In a nutshell, the strategic scenario can be described as process where a sender transmits a signal with the intention of sharing some piece of information where the receiver is then burdened to properly interpret the signal and choose a corresponding action. A strategic game is formulated with the preferences that both the sender and receiver have over the outcomes that result from signals sent and the actions taken.

One player, the *sender*, $\mathtt{s}$, has information that the other player, the *receiver*, does not. Player $\mathtt{s}$ has private knowledge of whether he is in a state (or is of type) $t \in T$, and when $\mathtt{s}$ is often interested in the actions taken by $\mathtt{r}$, the communication of $\mathtt{s}$'s private information becomes a game. The signalling game that ensues involves the process of $\mathtt{s}$ trying to guide the action $a \in A$ taken by $\mathtt{r}$ by means of *sending* a signal, or message, $m \in M$ that “don't have a pre-existing meaning” [38].

The set of strategies available to $\mathtt{s}$, S, is described by a function from states $t \in T$ to signals $m \in M$, $S : T \to M$. That is, $\mathtt{s}$'s choice of strategy is among a set of state–signal pairs. The set of strategies available to $\mathtt{r}$, R, is a function from signals $m \in M$ to actions $a \in A$, $R : M \to A$, resulting in signal–action pairs.

As an example, suppose $T = \{winter, summer\}$, $M = \{\text{“}brrr\text{”}, \text{“}phew\text{”}\}$, and

$A = \{open\ windows,\ make\ a\ fire\}$. The following tables describe $\mathtt{s}$'s and $\mathtt{r}$'s strategies, respectively:

	winter	*summer*
S_1	*"brrr"*	*"brrr"*
S_2	*"brrr"*	*"phew"*
S_3	*"phew"*	*"brrr"*
S_4	*"phew"*	*"phew"*

and

	"brrr"	*"phew"*
R_1	*open windows*	*open windows*
R_2	*open windows*	*make a fire*
R_3	*make a fire*	*open windows*
R_4	*make a fire*	*make a fire*

We can conclude that if $\mathtt{s}$ uses S_3 and $\mathtt{r}$ uses R_1 then they will be frozen in the winter and comfortably cool in the summer. Nevertheless, we know that something has gone wrong in communication here, because both $\mathtt{s}$ and $\mathtt{r}$ would prefer to not be popsicles in the winter. The combinations with desirable outcomes are S_2 and R_3, or S_3 and R_2.

The utilities for $\mathtt{s}$ and $\mathtt{r}$ are based on the actual state t, S and R, but they are not calculated in the same way: $U^*_{\mathtt{s}}$ is defined by $\mathtt{r}$'s response R to $\mathtt{s}$'s strategy S, whereas $U^*_{\mathtt{r}}$ must be measured as *expected* utility, because he has incomplete information regarding S's strategy; "it might be that the sender using strategy S sends in different states the same signal m" [38]. Nevertheless, we can construct a table with values for the outcomes in the above example, where the cells represent each (t, a) pair:

	open window	*make a fire*
winter	0,0	1,1
summer	1,1	0,0

A strategy profile, as usual, is indicated by the strategy played by the players $\mathtt{s}$ and $\mathtt{r}$: the pair $\langle S, R\rangle$. The Nash equilibrium in a signalling game is thus [38]:

Definition 5.1.1. *A strategy profile $\langle S, R\rangle$ forms a **Nash equilibrium** if and only if for all $t \in T$ the following two conditions are obeyed:*

$$(i)\ \neg\exists S' : U^*_{\mathtt{s}}(t, S, R) < U^*_{\mathtt{s}}(t, S', R)$$

$$(ii)\ \neg\exists R' : U^*_{\mathtt{r}}(t, S, R) < U^*_{\mathtt{r}}(t, S, R')$$

In many cases there are multiple Nash equilibria, which make it difficult to claim that it is responsible for establishing meaning. We can, however, refine the Nash equilibria with the concept of separating equilibria. A separating equilibrium ensures that, for instance, the sender does not use one signal for two states. Thus, "different messages are sent in different states such that there is a 1-1 correspondence between meanings and messages" [38]. One might argue that a separating equilibrium is a good characterisation of successful communication, but given that there can still be multiple separating equilibria, there must be a good reason to choose one over the other towards a unique solution. Lewis claims that one equilibria will be more likely since each player will expect the other to play that equilibrium strategy, where that expectation arises from linguistic convention. Consider the above summer and winter example: it has two separating equilibria $\langle S_2, R_3 \rangle$ and $\langle S_3, R_2 \rangle$, but there is no concrete reason for "brrr" being a conventional signal indicating that it is cold; picking one of those separating equilibria over the other as the convention is arbitrary.

Lewis defers to salience as a higher order explanation for choosing one equilibrium over the other; a salient equilibrium has some psychological or non-linguistic feature that sets it apart from the other separating equilibria. Besides this unsatisfactory explanation, Lewis' convention "makes a strong *rationality* assumption concerning the agents engaged in communication. Moreover, as for all equilibria in standard game theory, a lot of *common knowledge* is required; the rules of the game, the preferences involved, the strategies being taken ..., and the rationality of the players must all be common knowledge" [38].

Convention Justified by Evolutionary Stability

Van Rooij argues that evolutionary game theory is a more appropriate framework to explain stable linguistic conventions. It turns out that if separating equilibria are refined with the concept *evolutionary stable strategy*, the uniquely selected equilibrium is the equilibrium that Lewis intended to be a self-sustaining signalling convention. The improvement afforded by refining with ESS (and in general, by taking an evolutionary game theoretic perspective) is that we are no longer bound to stringent rationality assumptions.

The replicator dynamics in evolutionary game theory also solve the problem of the emergence of a signalling convention. Recall the an asymptotically stable equilibrium "is a solution path where a small fraction of the population starts playing a mutant strategy still converges to the stable point" [38]. Thus, the signalling convention is always the strategy that eventually all the members of a population adopt.

Conclusion

The multiple separating equilibria and the "salience" explanation for one becoming a convention over the other is, as van Rooij pointed out, a weak argument and requires pressing rationality assumptions that are difficult to justify given the limited interaction between players, imperfect information and uncertain intentions of the sender. Therefore, the fact that it is a benefit to theories such as pragmatics that evolutionary game theory purges rationality suggests that perhaps we ought to embrace a "type" of game theory without rationality. But how do we formalise a game theory without rationality? The next section addresses this query.

5.1.3 An Acceptable Way to Abandon Rationality

The arguments in this section justify finding an acceptable way to make abandoning rationality in a game theoretic setting acceptable under the right circumstances. We have examined three alternative views of rationality, which all have drawbacks. However, the literature on game theory and semantics has indicated that a variation that is less rationally demanding is welcome in at least the study of linguistics. The relationship, or bridge law, between information, preferences and actions *is* preserved in evolutionary game theory, but it is not "rationality" in the philosophical sense. In fact, it has been claimed many times in this thesis that in evolutionary game theory, strategies are reinterpreted as an expression of a gene. If this is true, then we can claim that the "space" between the player, his actions and preferences just disappears, as if the relationship replacing rationality is equality. If this is the case, may motivate us to let go of rationality as necessary component of game theory, for under the group dynamics setting of evolutionary game theory, we get the same result. If not a player is not"rational", then he cannot "exist". This suggests that there is I stated above that I would not claim that game theory should be void of a bridge law; what I claim is that one acceptable form of "bridge law" is by means of collapsing *strategy* and *player* into one term. The system is not void of a bridge law, but the concepts being bridged became one, also eliminating the *need* for a bridge law altogether. By doing this, we can also adjust the standard notion of preference to fit this evolutionary interpretation. The next section will fully describe an appealing logic where strategy and player become one formal term.

5.2 Logic

This section proposes two ways to think of evolutionary game theory in terms of logic. First, we look at the game matrix used to define the concept evolutionary stable strategy (ESS), observe a way to define it in modal logic, and see that it is not possible. Following this, a second attempt will be made using freedom, knowledge and preference

5.2.1 A Hybrid Logic for Evolutionary Stable Strategies

This section will propose an initial hybrid logic translation of ESS based on the relations freedom, knowledge and preference from chapter 4. There are some strange properties of this solution concept, however, that pose challenges to formulating a definition that properly expresses ESS. Thus, I will first describe some of those formal dilemmas, and then explain how a definition in terms of hybrid logic ought to remedy those solutions. Unfortunately, hybrid logic also has its drawbacks with regards to how accurately it represents what ESS is meant to express.

Formal Dilemmas

Suppose we attempt to define an ESS by means of the modalities for preference, $\langle \succeq_i \rangle$, freedom $\langle \approx_i \rangle$, and knowledge $\langle \sim_i \rangle$ and the possible intersection modalities. Recall a strategy σ is an evolutionary stable strategy if for all $\tau \neq \sigma$:

1. $u(\sigma, \sigma) \geq u(\tau, \sigma)$, which is the Nash equilibrium, and

2. If $u(\sigma,\sigma) = u(\tau,\sigma)$, then $u(\sigma,\tau) > u(\tau,\tau)$ which ensures that if a σ–player encounters a mutant τ–player, the strategy σ is still better than adopting the mutant strategy τ, so the mutant(s) will die off.

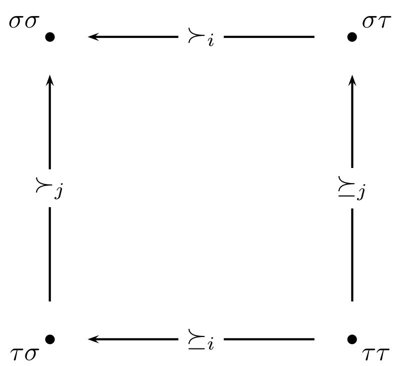


Figure 5.1: This figure depicts a 2–player game in which the arrows represent the preference requirements for ESS for both players i and j.

I will describe the procedure behind attempting to construct a modal logic definition of ESS, including observations that reveal the challenges that arise from doing so. The following observations shed light on these challenges:

(i.) The first condition is straightforward, for it requires that when both players play σ, it is a Nash equilibria. Thus,

$$\mathcal{M}, (\sigma,\sigma) \models \neg\langle \succ_i \cap \approx_i \rangle \top \text{ for all } i \in N$$

This expresses that "there is no preferred state accessible from (σ,σ) that is more preferred by i where his opponent plays a different strategy." Because this holds for both players, the current state is a Nash equilibrium.

(ii.) The second condition postulates that if there exists a state accessible from (σ,σ) that is equally preferred by i where his opponent plays a different strategy, then it must be the case that i (strictly) prefers (σ,τ) to (τ,τ). By simply applying the modalities, we get:

$$\mathcal{M}, (\sigma,\tau) \models \neg\langle \succeq_i \cap \approx_i \rangle \top \text{ for all } i \in N$$

(iii.) This formula can only be correct if the players have only two possible strategies σ and τ. If there is a third (or more) strategy ρ, then this formula implies that (σ,τ) must also be better than (σ,ρ) for any $\rho \in S$. In the case of a third strategy ρ, the second condition for σ to be an ESS would be that (σ,ρ)

is (strictly) preferred to (ρ, ρ).

(iv.) This is problematic given what ESS stands for: it represents the conditions under which a population is homogeneous and robust against mutations. That (τ, σ) is more preferred than (τ, ρ) gives us no information about the strength of the strategy τ.

(v.) In order to ensure that these preferences for i only run between (σ, x) and (x, x) (where x is some strategy available to i) and *not* between (σ, x) and (y, x) for some other strategy $y \neq x$, we should identify a specific relationship between (σ, x) and the states that are on the "diagonal" where both players are playing the same strategy and in their freedom relations.

(vi.) So far, there is no way to define the relationship between (σ, σ) and (τ, τ) for all $\tau \neq \sigma$. The logic does not seem to support the reference ability that is needed to pinpoint a state on the diagonal.

(vii.) This is likely not the only problematic feature of a modal logic definition of ESS. A problem that arises from at least the approach described in observations (i.) and (ii.) is that the two formulas have different points of reference. The first condition of ESS has as the point of reference (σ, σ) and the second condition has (σ, τ).

With the tools of hybrid logic, however, we can attempt another modal logic definition of ESS where we can pinpoint the states that are on the diagonal, and it can all be defined with (σ, σ) as the reference point.

Proposed Hybrid Logic Solution

The first proposal for a hybrid logic translation of the ESS is as follows:

Definition 5.2.1. *A strategy σ is an* ***evolutionary stable strategy*** *if*

1. *The first condition, which is that (σ, σ) must be a Nash equilibrium, will obviously be written in hybrid logic as:*

$$\mathcal{M}, (\sigma, \sigma) \models F_i \downarrow p \ P_i^{\succ} \downarrow q \ \neg(p \wedge q) \ \textit{for all } i \in N$$

2. *The second, which must have the same point of reference:*

$$\mathcal{M}, (\sigma, \sigma) \models K_i \downarrow r \ (q \wedge F_i \downarrow r P_i^{\succeq} \downarrow t \ \neg(r \wedge t)) \ \textit{for all } i \in N$$

This definition is accurate, but not intuitive. It still makes use of the notion of preference and freedom which actually should be reinterpreted properly into terms that do not insinuate that choice and preference are part of evolutionary game theory. Therefore, I propose the following definition as a more appropriate approach to evolutionary game theory.

5.2.2 Logical Reinterpretation of Classical Terms

There exists a mismatch between classical game theoretic terminology reinterpreted in evolutionary game theory and the *meaning* of evolutionary game theory. It is prudent at this point to re-evaluate whether the logic we have used up to this point really expresses what we intend it to. It has been stressed throughout this thesis that choice and rationality rejected by evolutionary game theory. The reinterpretation also implies that the player is significantly less powerful; a player is only a vessel for a strategy. However, the logical approaches to evolutionary game theoretic concepts are based entirely on logics invested in choice, knowledge and preference. It seems that there ought to be a more tailored way to speak of evolutionary game theory than that. The following observations gives a clue on how to the logic to reflect this issue.

> The treatment of rationality as preference consistency [...] allows us to assume that agents choose best responses [...]. How do [animals] accomplish these feats with their small minds and alien mentalities? The answer is that the *agent* is displaced by the *strategy* as the dynamic game theoretic unit. [23]

A logician can take Gintis' suggestion that "the agent is displaced by the strategy" to heart. Suppose we therefore reinterpret the set N of players i into the set $\mathbf{S} = \{\sigma, \tau, \rho, ...\}$ of players programmed with some strategy σ, τ, ρ, ...; that is, σ–players, τ–players, ρ–players, etc. This section will describe the logical reinterpretation of classical game theory.

Players to Strategy–Players For simplicity, we consider only two–strategy–player games; this is also consistent with the evolutionary interpretation which postulates that encounters are between two random members of a population. So, if in an evolutionary game the population is composed of σ- and τ–players then $\mathbf{S} = \{\sigma, \tau\}$ and an encounter of the players in a game is a strategy profile: some $\mathbf{s} \in \mathbf{S} \times \mathbf{S}$ such as $\mathbf{s} = (\sigma, \sigma)$ or $\mathbf{s}' = (\sigma, \tau)$. $\mathbf{S}$ successfully collapses a player and his strategy into one term.

Strategy Profiles to Pairings What we might otherwise call a strategy profile $\mathbf{s} \in \mathbf{S} \times \mathbf{S}$, we now call a pairing, because $\mathbf{s}$ will represent the encounter of two strategy–players. We saw in section 4.1.3 that [10] describes the semantics of the relations $\sim$ and $\approx$ with[3] σ as an outcome in a strategic game where $\sigma \models p_i^a$ expresses "player i plays action a in σ." We can also describe what happens in a pairing $\mathbf{s}$ in this way. The atomic proposition[4] e_σ expresses "an encounter involving a σ–player and some other strategy–player"

$\mathbf{s} \models e_\sigma$ if and only if $\mathbf{s} = (\sigma, \tau)$ for some $\tau \in \mathbf{S}$.

[3]Not to be confused with σ that has here been reinterpreted as a strategy unit.
[4]e is for enounter

Utility and Preference to Fitness As for the notion of utility or preference, evolutionary game theory replaces it with the notion of reproductive fitness, as indicated in chapter 3. We can thus also reinterpret the standard notions utility and preference into reproductive fitness. This ought to express that every pairing has a value for each player, where that value is the *amount* of offspring[5]. The amount of offspring for σ given the strategy–player he encounters is determined by a utility–like function, rep, from pairings to a value in $\mathbb{R}$, $\mathsf{rep}_\sigma : \mathbf{S} \times \mathbf{S} \to \mathbb{R}$.

Some pairings lead to higher fitness for a σ–player than other pairings, and therefore a σ–player "prefers" to meet certain kinds of players over others; this notion is defined in terms of fitness for a σ–player; that is, $\sigma \succeq_\sigma \tau$ expresses that a σ–player has better fitness from a pairing with another σ–player than a τ–player. Formally,

Definition 5.2.2. *A pairing with a τ–player is better for the fitness of a strategy σ than in a pairing with a τ'–player*

$$\tau \succeq_\sigma \tau' \textit{ if and only if } \mathsf{rep}_\sigma(\sigma, \tau) \geq \mathsf{rep}_\sigma(\sigma, \tau')$$

The relationship between fitness, $\succeq_\sigma$, and amount of replication is comparable to the relationship between preference and utility described in section 2.1.

The $\succeq_\sigma$ relation runs between alternatives for σ's opponent within a pairing **s** and not between actual pairings **s**, **s'**. However, the $\succeq_\sigma$ relation does induce a relationship between pairings. Because $\tau \succeq_\sigma \tau'$, the pairings $\mathbf{s} = (\sigma, \tau)$, $\mathbf{s}' = (\sigma, \tau')$ must be in what we might call the knowledge relation. The only varying factor is σ's opponent in each pairing, which is precisely what the knowledge relation does. Thus,

Fact 5.2.1. *If $\tau \succeq_\sigma \tau'$, then $(\sigma, \tau) \simeq_\sigma (\sigma, \tau')$*

If the consequence did not hold, then the $\succeq_\sigma$ could not hold.

A very appealing feature of $\succeq_\sigma$ is that it encodes the symmetric nature of the game. With respect to $\succeq_\sigma$ as well as $\succeq_\tau$, the pairings (σ, τ) and (τ, σ) are "equal." After all, the pairing is only meant to express that a σ player encounters a τ player; this is true no matter how it is listed in the strategy profile. [6]

Given the above definitions of $\succeq_\sigma$ we may define an evolutionary game model Γ^E.

Definition 5.2.3. *A model for an* ***evolutionary game*** *is:*

$$\Gamma^E = \langle \mathbf{S}, (\succeq_\sigma)_{\sigma \in \boldsymbol{S}} \rangle$$

[5]The 'winner' produces offspring proportional to the replication function. This means that if the fitness of **s** for σ is 2 and the fitness of **s** for τ is 4, then strategy σ has twice as much offspring as τ, which constitutes winning.

[6]$\succeq_\sigma$ is order invariant! This holds because $\mathsf{rep}_\sigma(\sigma, \tau) = \mathsf{rep}_\sigma(\tau, \sigma)$ and $\mathsf{rep}_\tau(\sigma, \tau) = \mathsf{rep}_\tau(\tau, \sigma)$. Thus, it must be the case that $(\sigma, \sigma) \succeq_\sigma (\sigma, \tau)$ *implies* $(\sigma, \sigma) \succeq_\sigma (\tau, \sigma)$. Moreover, there is never a $\succeq_\sigma$ relation between worlds such as (τ, σ) and (τ, τ) because only the τ–player can prefer one of these strategy profiles over the other, since one has no σ player

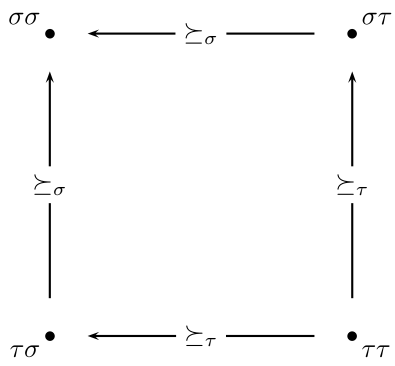


Figure 5.2: Some relations $\succeq_\sigma$, $\succeq_\tau$ are marked in this arbitrary game.

With only relations, we can characterize ESS as expressing that $\sigma \in \mathbf{S}$ is an ESS if and only if for all $\tau \neq \sigma \in \mathbf{S}$: $\sigma \succeq_\sigma \tau$ and $\sigma \succ_\tau \tau$

This encodes that it is best for a σ–player to play against a σ–player, because the σ–player has a higher fitness in that case than if he were to play against a τ–player. Moreover, a τ–player must also have a higher fitness from an encounter with a σ–player than with another τ–player. This ensures that, as ESS requires, σ–players will eventually out-replicate τ–players.

However, for this to fit the picture of modal logic, we must consider the model and the point of reference. Thus the proper definition of ESS is:

Definition 5.2.4. $\sigma \in \mathbf{S}$ *is an* ***evolutionary stable strategy*** *if and only if:*

$$\mathcal{M}, (\sigma, \sigma) \models \succ_\sigma \downarrow p(\neg p \wedge \succ_\tau \downarrow qA\neg q)$$

There are a few significant benefits of defining ESS in this way. This hybrid logic formulation encodes the grid structure as well as the 2-player nature of encounters in evolutionary game theory. This is because $(\sigma, \tau) \simeq_\tau \tau$ that is also related by $\succeq_\tau$ to (τ, σ) *and* (σ, τ). Moreover, n accordance with how evolutionary game theory is "constructed," the $\succeq_\sigma$ relation successfully separates the notions of choice and preference from the solution concept. Another benefit is that it encodes that the game must be symmetric, because a relation $\succeq_\sigma$ cannot exist between a world where there is a match including a σ–player and a world where that does not include a σ–player.

5.3 Conclusion

The goal of this chapter was two-fold: it tackled the problems that arose from the two factors in evolutionary game theory that were reinterpreted from classical game theory. Those two factors were, as described earlier, that natural selection replaced rationality and that fitness replaced utility. The abandonment of rationality alone posed problems for understanding evolutionary game theory as a theory related to classical game theory. However, even though attempts to "save" rationality in evolutionary game theory, there are good reasons to suggest that *not* having rationality is manageable, even in when attempting to describe evolutionary game theoretic concepts in terms of logic involving all kinds of epistemic factors like choice, preference and action.

This brings us to the end of the thesis. In the following section, I will describe what this thesis has accomplished and further steps that can be taken in evolutionary game logics.

Chapter 6

Conclusion

The initial query in this thesis was whether a evolutionary game logic could be established given the established classical game logics. Classical game logics are often modal logics with epistemic modalities, because classical game theory has an inherently epistemic character; many concepts and solutions revolve around players' knowledge, choices and preferences. Logicians thus took the initiative to explore how epistemic logic could logically describe game theory. From those investigations, logicians formulated elegant and concise formulas representing the most central solution concepts of classical game theory. We saw, for instance, that Nash equilibrium could be expressed by the following simple formula:

$$\mathcal{M}, \sigma \models \neg\langle\succ_i\rangle\varphi$$

But evolutionary game theory has not been connected to logic like classical game theory has, despite the fact that evolutionary game theory seems like a rather small step away from classical game theory. Although evolutionary game theory has adopted classical game theory's formal set-up (with players, strategies, preferences, matrices among others), the crucial difference between the two is the reinterpretation of the terminology. This is responsible for many obstacles in introducing logic to evolutionary game theory. A goal of this thesis is thus to explore the possibilities and challenges of formulating an evolutionary game logic.

To this end, chapters 2 and 3 introduced the main ideas and results of classical and evolutionary game theory, respectively.

Chapter 2 introduced the basic machinery used to model strategic interaction and built the language for the rest of the thesis. Classical game theory was shown to be composed of two main forms, the strategic form game and the extensive form game. The terminology for strategic form games, which would be the main way to talk about all things game theoretical throughout this thesis, was initially introduced to the reader. The game matrix and solution concepts were described next, emphasizing the importance of one result in particular, the Nash equilibrium. Last, this section described the process of iterated elimination of dominated strategies (IEDS). This process would be a central example in section

4.1, because it demonstrates how the dynamic process of declaring rationality allows the players to reason towards a solution of a game.

Following the strategic form game, its counterpart the extensive form game was introduced. This section began by describing the terminology that accounts for the sequential nature of this game form. The game tree further expressed the elaborate features of a turn-taking game. Although, the extensive game is an obviously different way of formalising a game than the strategic game, the solution concepts, such as Nash equilibrium still hold, and are defined in a similar way. However, Nash equilibrium does not reflect the nature of an extensive form game; therefore, we have the additional solution concept of subgame perfect equilibrium, which does. Normalisation was shown to be a procedure on extensive form games relevant to the discussion on logic and extensive form games. It was necessary to motivate why relations like freedom ought to be possible in extensive form games. Describing the procedure of backward induction was also crucial to the results in this thesis; as we saw in section 4.2, it was (as IEDS was in 4.1) the "pilot" example of how players reasoned towards an outcome of the game.

Future Work 1. I mention here that IEDS and BI are both processes in a classical game involving the rational reasoning towards the solution of a game. Of course, IEDS is a process for strategic form games, whereas BI is a process for extensive form games. It was not discussed in this thesis how these procedures are related to each other; especially in light of the other comparisons I made with factors of classical game theory across the strategic and extensive form (such as the manifestations of the relationships freedom, knowledge and preference in both forms). It is thus a suggestion for further development that both procedures are analysed with the following questions in mind [6]: Does the BI procedure on an extensive form game select the same outcomes as the IEDS procedure on the normalized version of the extensive form game? Does the process itself eliminate outcomes or strategies in the same order? Or is the only similarity the BI solution and the strategy profile surviving IEDS?

Chapter 3 introduced evolutionary game theory as the second main sort of game theory relevant to this thesis. The history behind evolutionary game theory is unique in that it stems from evolutionary biology, and its theoretical roots are in Darwin's theory of natural selection. Following this, it was made explicit that evolutionary game theory, which utilizes the same machinery as classical game theory, reinterprets those concepts within the theoretical framework of natural selection and evolution. This chapter stressed that the reinterpretation of game theoretic concepts were incompatible with the crux of classical game theory: rationality. It seems that in many accounts of evolutionary game theory, the abandonment of rationality in evolutionary game theory was acknowledged, but not seen as a concern or as something to act upon. I speculate that this is because the reinterpretations generally, do not cause many problems. But as we saw in the attempt at a logic for evolutionary game theory, the change in interpretation of rationality as well as the other terms created obstacles for defining concepts such as ESS. This section continued with a description of ESS, and then continued with a brief discussion of replicator dynamics. As mentioned in the conclusion of chapter 3, this thesis does not address the replicator dynamics side of evolutionary game theory in conjunction with logic. This reflects the fact that I wish to remind the reader of the value of strategic form games. In chapter 4, the strategic form game is decorated with new relations and a new logic (hybrid logic) to characterize its workings. After all, "the simple matrix pictures that one sees in a beginner's text on game theory are already models for quite sophisticated logics of action, knowledge and preference" [10]. The replicator dynamics are of course appealing to the modal logician, because it entails a model that changes over time. This connects clearly with modal logics; especially with temporal logics and topological logics [30]. Investigating a logic for replicator dynamics is complex, the replicator dynamics describes a model that is infinite and continuous (instead of discrete).

Future Work 2. Thus, I suggest as future work a similar approach to logic of evolutionary game theory, but with an emphasis on replicator dynamics.

Chapter 4 is the part of this thesis that introduces and describes already existing logics for classical games. The existing logics described in this chapter recognize that classical game theory, in which agents, interaction, preferences, decisions, deliberation and knowledge play leading roles, is an ideal fit for (modal) logic. In particular, there are (dynamic) epistemic logics that elegantly describe the inherently epistemic concepts of classical game theory (such as knowledge of the opponent).

In 4.1 some solid examples of existing modal logics for strategic form games were described; this included Bonanno [13], van Benthem [2], van Benthem's theory of play [10] [4], and last, a new perspective: a hybrid logic based on relations between outcomes in a game. These relations, freedom F_i or $\approx_i$, knowledge, K_i or $\sim_i$, and preference P_i or $\succeq_i$ were described in detail. These relations described games on their own or as intersection modalities, which also gave of elegant definitions of game theoretic concepts such as Nash equilibrium:

$$\mathcal{M}, \sigma \models \neg\langle \approx_i \cap \succ_i \rangle \top$$

This chapter continued to demonstrate how Seligman used these relations as modalities and applied them to hybrid logic. By means of hybrid logic's binder operator $\downarrow$, it was possible to label the worlds that were reachable by means of a particular relation.

Future Work 3. This thesis focused on the use of the *binder* from hybrid logic, but chapter 4 also introduced the satisfaction operator. This tool is likely to be applicable and valuable in the hybrid logic formalisation of game theoretic concepts.

In this section I addressed a minor mistake in Seligman's proposed hybrid logic definition of strictly dominated strategies, proposed hybrid logic definitions for van Benthem's concepts of weak and strong rationality, and proposed a way to describe the relationship between weak rationality and IEDS.

Part 4.2 addresses established modal logics in extensive form games. For consistency, I briefly outlined how Bonanno approached extensive form games. We could immediately see that as IEDS had been the example for rational deliberation, backward induction will be that for extensive form games. The theory of play took a more in-depth look at modal logics for extensive form games . This section focused on two specific topics within van Benthem et al.'s theory of play: the notion of extensive form games as process models, and the feature of a player's strategic powers in a game. These topics are relevant for this thesis, because my account of freedom in extensive form games is strongly related to both topics. Following this, we observed a proposal for a new definition of freedom for extensive form games, because the way we defined freedom in the discussion of strategic form games does not fit extensive form games. The rest of section 4.2 addresses how my formulation of freedom is related to the established modal logics for extensive form games as described by van Benthem.

Chapter 5 addressed the two crucial differences between classical and evolutionary game theory: the renouncement of rationality and the reinterpretation of classical game theoretic terminology. Thus this section examined both differences separately and evaluated the consequences. We first saw that by even by taking an alternative perspectives on rationality, there was no way to "fit" rationality into a theory of evolutionary games. took hybrid logic as a framework equipped to describe evolutionary game theoretic concepts.

Future Work 4. The debate of rationality's role in evolutionary game theory considered three alternative views of rationality, but there are, in fact, many different approaches to rationality. It would be valuable to the study of evolutionary game theory to thoroughly investigate a possible role for rationality. Eric Pacuit and Olivier Roy, in their upcoming text *Interactive Rationality*, which will "highlight the foundational/philosophical issues that are coming up in contemporary 'interactive epistemology,' an emerging field at the intersection of game theory, logic, computer science and philosophy" [36].

The section 5.1.2. takes a brief look at an application of evolutionary game theory to pragmatics in natural language. Besides observing how evolutionary game theory can be applied to a topic that is related to logic, language, this section also demonstrated that under some circumstances, the less dependence on rationality, the better. Section 5.1 thus concluded with the claim that abandoning rationality entirely is technically acceptable under given the interpretation of the conclusions that were drawn in this section.

Following the rationality discussion, section 5.2 demonstrated that the manner in which it is acceptable to abandon rationality entails a great suggestion for how to approach devising a logic in which concepts of evolutionary game theory were properly expressed. However, this section started with a motivation for devising an amended logic in the first place. It thus demonstrated some formal difficulties that arose from attempting to define ESS by means of "standard" modalities. Following this, a hybrid logic definition based on Seligman's relations as modalities was demonstrated, which did result in an acceptable definition for ESS. However, it would have been weak to conclude this thesis with a definition that is defined by terms that I have argued do not apply to evolutionary game theory (choice, knowledge, preference are all not compatible with the abandonment of rationality). Therefore, I suggest an alternative logic that defines ESS in a more theoretically consistent manner. Therefore, we resulted with another elegant formula, like the ones described above, that defines the main static solution concept in evolutionary game theory:

$$\mathcal{M}, (\sigma, \sigma) \models \succ_{\sigma} \downarrow p(\neg p \wedge \succ_{\tau} \downarrow qA\neg q)$$

I have mentioned some directions for future work throughout this conclusion, but given that the topic of game theory and logics is broad, there are many directions of further study; too many to mention. The following two suggestions ought to be addressed since the topics were omitted from this discussion:

Future Work 5. Fixed point logics are relevant to backward induction and IEDS, and it is also strongly related to dynamical systems. I do not address fixed point logics at in this thesis, but a deeper investigation will certainly be valuable to the topic.

Future Work 6. Hybrid logic, although a desirable and expressive logic, has its drawbacks. The complexity becomes greater, especially with the interaction of so many relations. This certainly warrants an investigation.

The goal of this thesis was make an initial fomrulation towards a logic of evolutionary game theory, bringing it into the game logics debate. This discussion, therefore, finds its place in the *logic of games* side of the game–logics interface described by Johan van Benthem in *Logic and Games.* I can conclude that I have shown that the evolutionary stability concepts can be understood in terms of hybrid logic, and that despite the challenges that arise from the abandonment of rationality and the reinterpretation of terms, some interesting and valuable results were still viable.

Bibliography

[1] Alexandru Baltag, Sonja Smets, and Jonathan Zvesper. Keep hoping for rationality: a solution to the backward induction paradox. *Synthese*, 169(2):301–333, 2009.

[2] Johan van Benthem. Rational dynamics and epistemic logic in games. *International Game Theory Review*, 9(1):13, 2007.

[3] Johan van Benthem. 'Games that make sense': logic, language, and multi-agent interaction. *New perspectives on games and interaction*, 4:197–209, 2008.

[4] Johan van Benthem. *Logic in Games.* Unpublished Manuscript, 2012.

[5] Johan van Benthem. *Modal Logic for Open Minds.* Center for the Study of Language and Information, 2012.

[6] Johan van Benthem. Personal communications. 2012.

[7] Johan van Benthem and Jan van Eijck. Game theory, logic and rational choice. *Discourses on Social Software*, 5:123, 2010.

[8] Johan van Benthem and Patrick Girard and Olivier Roy. Everything else being equal: A modal logic approach to ceteris paribus preferences. *Journal of Philosophical Logic*, 2008.

[9] Johan van Benthem and Sieuwert van Otterloo and Olivier Roy. Preference logic, conditionals and solution concepts in games. In *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg.* Uppsala Philosophical Studies 53.

[10] Johan van Benthem, Eric Pacuit and Olivier Roy. Toward a theory of play: A logical perspective on games and interaction. *Games*, 2(1):52–86, 2011.

[11] Kenneth Binmore. *Rational decisions.* Princeton University Press, 2009.

[12] Patrick Blackburn and Maarten Marx. Lectures on hybrid logic. First North American Summer School in Logic, Language, and Information, Stanford, 2002.

[13] Giacomo Bonanno. Modal logic and game theory: two alternative approaches. *Risk Decision and Policy*, 7(3):309–324, 2002.

[14] Adam Brandenburger. Game trees. Technical report, New York University, 2007.

[15] Bart J. Buter. Dynamic programming for extensive form games with imperfect information. Master's thesis, Universiteit van Amsterdam, 2012.

[16] Balder ten Cate and Massimo Franceschet. On the complexity of hybrid logics with binders. In *Computer Science Logic*, pages 339–354. Springer, 2005.

[17] Vincent P. Crawford. Nash equilibrium and evolutionary stability in large and finite populations. *Annales d'Economie et de Statistique*, (25/26), 1992.

[18] Charles Darwin. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* John Murray, London, 1859.

[19] David Easley and Jon Kleinberg. *Networks, crowds, and markets.* Cambridge University Press, 2010.

[20] Shih fen Cheng, Daniel M. Reeves, Yevgeniy Vorobeychik, and Michael P. Wellman. Notes on equilibria in symmetric games. In *In Proceedings of the 6th International Workshop On Game Theoretic And Decision Theoretic Agents (GTDT*, pages 71–78, 2004.

[21] Melvin Fitting. Reasoning about games. *Studia logica*, 99(1-3):143–169, 2011.

[22] Michael Franke. Free choice from iterated best response. In *Proceedings of the 17th Amsterdam colloquium conference on Logic, language and meaning*, pages 295–304. Springer, 2009.

[23] Herbert Gintis. *Game theory evolving: A problem-centered introduction to modeling strategic behavior.* Princeton University Press, 2000.

[24] Herbert Gintis. *The bounds of reason: game theory and the unification of the behavioral sciences.* Princeton University Press, 2009.

[25] Jaakko Hintikka and Gabriel Sandu. Game theoretic semantics. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*, chapter 6. 1997.

[26] Wiebe van der Hoek and Marc Pauly. Modal logic for games and information. In Blackburn et al., editor, *Handbook of Modal Logic*, chapter 20. Elsevier, 2007.

[27] Josef Hofbauer and Karl Sigmund. *Evolutionary Games and Population Dynamics.* Cambridge University Press, 1998.

[28] Gerhard Jäger. Evolutionary game theory for linguists. a primer, 2004.

[29] Gerhard Jäger. *Game Theory and Linguistic Meaning*, chapter Game dynamics connects semantics and pragmatics, pages 89–102. Elsevier, 2007.

[30] Philip Kremer and Grigori Mints. Dynamic topological logic. *Annals of Pure and Applied Logic*, 131:133–158, 2005.

[31] David Lewis. *Convention*. Harvard University Press, 1969.

[32] Markus M. Mobius. Lecture IV: The existence of Nash equilibrium. Technical report, Harvard Univeristy, 2008.

[33] John Nash. Non-cooperative games. *The Annals of Mathematics*, 54(2):286 – 295, September 1951.

[34] von Neumann, John and Morgenstern, Oskar. *Theory of games and economic behavior*. Princeton University Press, 1953.

[35] Martin Osborne and Ariel Rubinstein. *A course in game theory*. The MIT press, 1994.

[36] Eric Pacuit and Olivier Roy. Interactive rationality. Unpublished Manuscript.

[37] Andrés Perea. *Rationality in Extensive Form Games*. Kluwer Academic Publishers, 2001.

[38] Robert van Rooij. Evolution of conventional meaning and conversational principles. *Synthese*, 139(2):331–366, 2004.

[39] Robert van Rooij. Signalling games select Horn strategies. In *Proceedings of the sixth Annual Meeting of the Geselschaft fur Semantik*, volume 27, pages 493–527. Springer, 2004.

[40] Jeremy Seligman. Hybrid logic for analyzing games. In *Lecture at the Workshop Door to Logic, Beijing May*, volume 2010, 2010.

[41] John Maynard Smith. Evolutionary game theory. *Physica*, 22 D:43 – 49, 1986.

[42] John Maynard Smith and George Price. The logic of animal conflict. *Nature*, 246, 1973.

[43] Jorgen W. Weibull. *Evolutionary Game Theory*. MIT press, 1997.